(12)

ADA047900

A PROXY APPROACH TO MULTI-ATTRIBUTE DECISION MAKING

STANFORD UNIVERSITY

Kenneth R. Oppenheimer

ADVANCED DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

RESEARCH REPORT No. EES-DA-77-3 ✓

6

A PROXY APPROACH TO MULTI-ATTRIBUTE DECISION MAKING.

by

10 Kenneth R. Oppenheimer

9 Doctoral thesis,

DECISION ANALYSIS PROGRAM

Professor Ronald A. Howard
Principal Investigator

Sponsored by

Defense Advanced Research Projects Agency
ARPA Order No. 3052

Under Subcontract from

Decisions and Designs, Incorporated

and

National Science Foundation
NSF Grant GK-36491

15 NØØ14-76-C-ØØ74, ARPA Order-3Ø52

11 October 1977

12 125p.

DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS
Stanford University
Stanford, California 94305

SUMMARY

In many decision problems, the possible outcomes have several important dimensions of value. To identify the optimal alternative, the decision analyst must assess the decision maker's preferences over these multi-attribute outcomes. Two rival procedures for solving multi-attribute preference problems currently exist. These two procedures, global preference modeling and local preference modeling, each have advantages and disadvantages. This dissertation combines these two rival procedures in a new approach to multi-attribute decision making. This new combined method, called the proxy approach, uses the advantages of one technique to overcome the disadvantages of the other.

Global preference modeling procedures use normative assumptions together with a few assessments to construct a single function, in the large, ordering preferences over all outcomes. These global functions are mathematically simple and convenient, but they are very restrictive. The assumptions from which they are derived are reasonable locally, in the small, but when assumed globally, in the large, they often produce functions not truly representing the decision maker's preferences.

Local procedures provide an alternative approach that avoids restrictive assumptions. Instead of constructing a single preference function in the large, these procedures build a sequence of local preference models, in the small, each generating a trial solution. Each trial solution is better than its predecessor, so the trial sequence eventually reaches the optimum. Currently existing local procedures use successive linear approximations; these linear functions are poor preference models, so the iterative procedure is slow and inefficient. Since each iteration requires a time-consuming interaction with the decision maker, the slowly converging procedure is not practical.

This dissertation combines the desirable features of the global and local techniques in a new improved method. The normatively motivated preference models of the global procedure are incorporated as proxy functions in a local procedure. These proxies are better models of the true objective than are the linear approximations, so the resulting

trial sequence reaches the optimum much faster. The new proxy approach yields rapid convergence without restrictive assumptions.

After the theoretical aspects of the proxy approach are developed, the new algorithm is applied to a curriculum planning problem. This practical application was successful; the decision maker, previously unfamiliar with decision analysis, was able to provide the asssessments required at each iteration. With the help of various consistency tests, the tradeoff assessments generated trial solutions that converged rapidly to the optimal solution. Numerous insights into the interactive use of the algorithm were gained from this practical application.

CONTENTS

FIGURES

Figure

TABLES

Table

ACKNOWLEDGEMENTS

I take pleasure in acknowledging the advice and encouragement of Professors Ronald A. Howard and Richard D. Smallwood, under whose direction this dissertation was completed. Their support and guidance were invaluable throughout my tenure as a graduate student.

I owe a special debt of gratidude to Professor Richard Smallwood for arranging my research internship in the Analysis Research Group at the Xerox Palo Alto Research Center. This dissertation was written during my two-year internship at Xerox PARC. My friends and colleagues at Stanford and at Xerox provided important moral support throughout my graduate program. I would especially like to thank Dr. Shmuel Oren and Dr. Peter Morris for the time they spent in numerous discussions about this research. Thanks are also due Professor William K. Linvill and Dr. James E. Matheson for reviewing the manuscript and offering many useful comments.

Above all, I am forever grateful to my parents for their contributions throughout my education, and to my wife Bicky for her constant encouragement and patience.

Xerox Corporation provided the financial support and the stimulating research environment for this dissertation. My research internship at Xerox PARC made my doctoral program at Stanford an especially enjoyable and rewarding experience.

Supervision and publication costs were also partially supported by the Department of Defense Advanced Research Projects Agency and by the National Science Foundation.

CHAPTER I

INTRODUCTION

1.1 Multi-Attribute Decision Analysis

Decision analysis is a normative methodology for identifying the best alternative in a choice situation. In the first step of an analysis, the analyst models the relationships among the actions the decision maker can choose and the resulting outcomes he will face. All important variables influencing the outcome must be included. In the second step, the analyst assesses the decision maker's preferences over the outcome variables so the optimal decision can be selected. Figure 1.1, taken from Howard [16], illustrates this paradigm for the individual decision maker.

The decision variables $d_1, d_2, \dots, d_m$ represent the choices available to the decision maker. The state variables $s_1, s_2, \dots, s_n$ represent the environmental factors affecting the outcome. Information about the uncertain state variables is encoded in a joint probability distribution $\{s_1, s_2, \dots, s_n|\varepsilon\}$, where ε represents the decision maker's state of information. The state variables and decision variables interact through the system model to produce the outcome lottery $\{\underline{x}|\underline{d},\varepsilon\}$, a function of the decision vector $\underline{d}$. Each possible outcome has a unique setting of the x_i's in the multi-attribute outcome vector $\underline{x}$. The decision maker's preference ordering over these multi-attribute outcomes is encoded in a mathematical utility function that provides a ranking of the alternatives so the decision producing the highest expected utility $\langle u|\underline{d},\varepsilon\rangle$ can be selected.

Techniques for handling the single attribute problem are well developed and have been applied successfully in numerous cases [5],[17]. The multi-attribute problem, however, is much more difficult and still poses a formidable challenge to decision analysts. In this dissertation, I try to develop an improved procedure for solving the multi-attribute problem for the individual decision maker. I take the state variables, decision variables, and system model as given, and address the problem of selecting the optimal decision. This dissertation is an attempt to find a practical procedure to overcome several disadvantages of currently-used multi-attribute techniques, with the

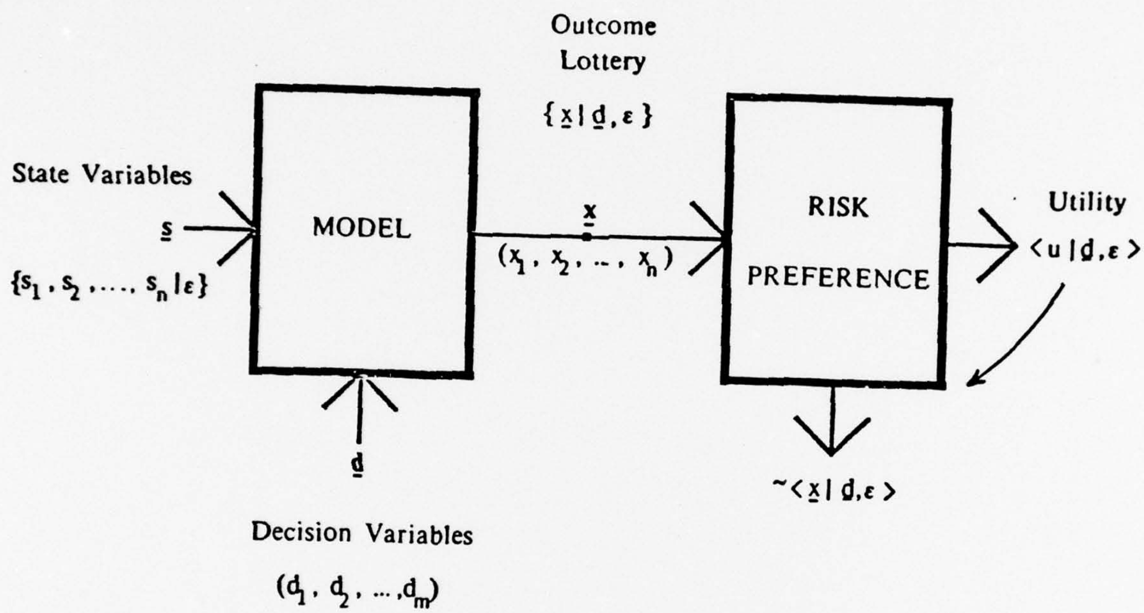


Figure 1.1 Paradigm for Multi-attribute Decision Analysis

hope of enabling decision analysts to deal more effectively with complex decisions.

1.2 Global vs. Local Procedures

Two different approaches are available, at least theoretically, to solve any multi-attribute problem. The first is an exhaustive comparison of alternatives; with this approach, the analyst assesses preferences at every possible outcome, thereby obtaining directly the complete preference ordering. This exhaustive search requires a large, often infinite number of assessments; it has little or no practical application. The second approach is a modeling technique; the analyst uses behavioral assumptions together with a few assessments to model preferences analytically. Decision analysts use this modeling approach since it provides efficient procedures for analyzing multi-attribute problems.

I divide these preference modeling procedures into two broad categories: global procedures and local procedures. Global procedures construct a single preference function ordering all outcomes; local procedures do not construct such a function. When using a global procedure, the analyst assesses preferences at a few outcomes and makes normative assumptions that uniquely specify the preference ordering at all other outcomes. These assumptions restrict the preference function to specific families of curves. Once the family is selected, a small number of assessments determine the free parameters. The resulting preference function applies over the entire decision region, hence the name global preference function.

These global functions are mathematically simple and convenient, but they have disadvantages. The assumptions from which the specific functional forms are derived are very strong conditions. They are reasonable in local regions, in the small, but when assumed globally, in the large, they are very restrictive. There may be instances in which a function fit from an assessment in the small adequately represents preferences in the large; in these instances the global procedures should be used. Generally, however, these global procedures force the decision maker to fit a function not truly representing his preferences. If the decision maker has a non-additive preference structure, the problem is particularly acute since all commonly used preference functions have additive

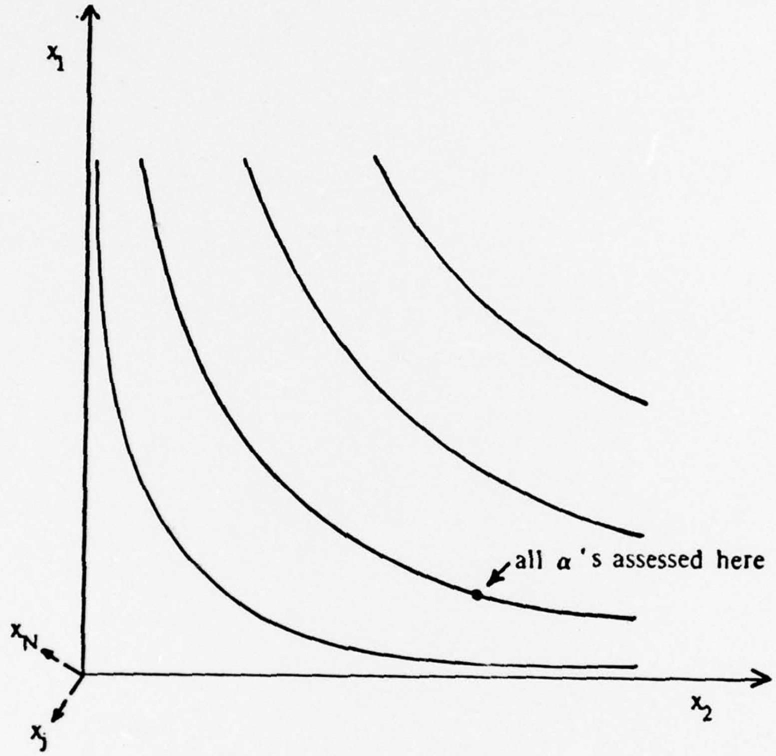
deterministic forms. The Cobb-Douglass example in Figure 1.2 shows the restrictiveness of global parameterization from local assessment. The parameters assessed at one point must characterize behavior over the entire region. Even with consistency checking at a second point, one set of parameters must ultimately hold everywhere.

These disadvantages of global preference modeling motivate the search for an alternative approach. Local procedures, the second category of preference modeling techniques, provide this alternative. If we are willing to construct a series of local preference models, in the small, we can develop an iterative procedure that scans the set of alternatives to locate the optimum. At each iteration of this procedure, the local preference model provides a trial solution. Each trial solution is better than its predecessor, so the sequence eventually reaches the optimum. This iterative technique avoids the strong restrictions; it never specifies a global preference function. Local procedures, however, require more assessments than global procedures, often too many assessments to be practical. This dissertation is an attempt to find a local procedure to solve multi-attribute problems, avoiding the restrictions of global preference modeling, and at the same time, requiring only a reasonably small number of assessments.

1.3 Outline of Thesis

Dean Boyd [4] made the first attempt to solve this basic problem cast in a decision analysis framework. He developed a procedure using local tradeoff assessments to parametrize local linear approximations of the true preference function. His procedure avoids global restrictions, but requires too many assessments to be practical. I review Boyd's thesis and other related literature in Chapter II.

My major contribution is a merger of the global and local procedures. I develop a new algorithm that incorporates the desirable features of both techniques; it uses the *normatively motivated* preference models of the global procedure as *proxy* functions in a local procedure. These proxies are better models of the true preference function. Therefore, the sequence of trial solutions they generate reaches the optimum much faster. This new proxy algorithm uses the advantages of one technique to overcome the



Cobb-Douglass: $V(\underline{x}) = x_1^{\alpha_1} \, x_2^{\alpha_2} \ldots x_N^{\alpha_N}$

Figure 1.2 Global Modeling from Local Assessment

disadvantages of the other; the result is a combined procedure that improves both original techniques. Chapter III presents the theoretical aspects of my methodology for decision making under certainty. It includes several applications of optimization theorems that help solve the multi-attribute decision interpreted as a resource allocation problem. Chapter III also includes tests of assessment consistency that keep the trial sequence from going astray.

Chapter IV compares the new algorithm to the old. Convergence comparisons show the new algorithm is much faster; it requires few enough assessments to be practical for decision making under certainty.

In Chapter V, I examine the proxy approach under uncertainty. Theoretical aspects of decision making under uncertainty present major obstacles. Consequently, the proxy approach in its current form is not useful for problems in which uncertainty plays a major role.

The true practical test of any theoretical decision-making procedure is a real problem. Chapter VI describes the application of the proxy approach to a curriculum planning problem of a small private school in San Jose, California. This practical application was successful; it provided numerous insights into the interactive use of the procedure.

Chapter VII summarizes the key results of the thesis and includes suggestions for future research.

CHAPTER II

RELATED LITERATURE

2.1 Axioms of Deterministic and Risk Preference

Most treatments of decision theory begin with the assumption that deterministic preferences can be represented by a set of three binary relations defined over all outcomes $\underline{x}$. The three relations are:

i. *Strict Preference* $\succ$

$\underline{x}^1 \succ \underline{x}^2$ if $\underline{x}^1$ is strictly preferred to $\underline{x}^2$

ii. *Indifference* $\sim$

$\underline{x}^1 \sim \underline{x}^2$ if $\underline{x}^1$ is indifferent to $\underline{x}^2$

iii. *Weak Preference* $\succsim$

$\underline{x}^1 \succsim \underline{x}^2$ if $\underline{x}^1 \succ \underline{x}^2$ or $\underline{x}^1 \sim \underline{x}^2$

Four axioms governing the decision maker's behavior under certainty are listed below. All are quite standard and can be found in most developments of preference theory [23].

Axiom 2.1. *Weak ordering.* The relation $\succsim$ is transitive and connected; $\succsim$ is transitive if $\underline{x}^1 \succsim \underline{x}^2$ and $\underline{x}^2 \succsim \underline{x}^3$ imply $\underline{x}^1 \succsim \underline{x}^3$; $\succsim$ is connected if $\underline{x}^1 \succsim \underline{x}^2$ or $\underline{x}^2 \succsim \underline{x}^1$ for all $\underline{x}^1$ and $\underline{x}^2$. This axiom prevents the decision maker from being a "money pump". A violation of transitivity implies a willingness to pay to accomplish nothing.

Axiom 2.2. *Continuity.* If $\underline{x}^1 \succsim \underline{x}^2$ and $\underline{x}^2 \succsim \underline{x}^3$, then there is a real number c, $0 \leq c \leq 1$, such that $c\underline{x}^1 + (1-c)\underline{x}^3 \sim \underline{x}^2$. This axiom indicates the decision maker is willing to make tradeoffs.

Axiom 2.3. *Nonsatiety.* If $x_i^1 \geq x_i^2$ for all i and $x_j^1 > x_j^2$ for some j, then $\underline{x}^1 \succ \underline{x}^2$. This axiom means the individual prefers more to less of each attribute. The outcome attributes must be modeled so each x_i is a desirable good.

Luce and Suppes [23] proved that Axioms 2.1-2.3 guarantee the existence of a

continuous real-valued deterministic preference function $V(\underline{x})$ such that $V(\underline{x}^1) \geq V(\underline{x}^2)$ *if and only if* $\underline{x}^1 \gtrsim \underline{x}^2$. *The function* $V(\underline{x})$ *is unique up to a monotonic* increasing transformation.

Axiom 2.4. *Decreasing Marginal Rates of Substitution.* For any x_i^1 and x_j^1, the amount of x_i traded for each additional Δx_j decreases as x_j increases. Assuming differentiability, $\partial[-(dx_i/dx_j)|_{dV=0}]/\partial x_j < 0$. This axiom states that the decision maker becomes less sensitive to incremental changes Δx_j as his amount of x_j increases.

These four axioms provide the foundation for mathematical preference structures for deterministic decision making. In the remainder of this thesis, we refer to them as the deterministic preference axioms. Analogous results guaranteeing the existence of a real-valued risk preference function for decision making under uncertainty were pioneered by von Neumann and Morgenstern [27] and later revised by Savage [31]. A convenient form of the Savage axioms, taken from Howard [18], is listed below. In the reference lottery $[p_x, x; p_y, y; p_z, z]$, p_x, p_y, and p_z are the probabilities of prizes x, y, and z, respectively.

Axiom 2.5 *Orderability.* For all x and y, either $x > y$, $x \sim y$, or $x < y$.

Axiom 2.6. *Continuity.* If $x > y > z$, then there is a real number p, $0 \leq p \leq 1$, such that $[1, y] \sim [p, x; (1-p), z]$. The quantity y is called the certain equivalent of the lottery.

Axiom 2.7. *Substitutability.* A lottery and its certain equivalent are interchangeable with no change in preferences.

Axiom 2.8. *Monotonicity.* If $x > y$, then $[p, x; (1-p), y] > [p', x; (1-p'), y]$ if and only if $p > p'$.

Axiom 2.9 *Decomposability.* $[p, \{q, x; (1-q), y\}; (1-p), y] \sim [pq, x; (1-pq), y]$.

In the remainder of this thesis, we refer to Axioms 2.5-2.9 as the risk preference axioms.

2.2 State-of-the-Art Techniques

The literature on multi-attribute utility theory is voluminous; it includes contributions from economists, decision analysts, and mathematical psychologists. In this thesis, I review only those techniques that rest upon a normative foundation and relate directly to my results. Within the decision analysis profession, there are currently two different approaches to assessing multi-attribute utility functions. One approach is associated with the Stanford Research Institute and Stanford University and the other with Harvard and M.I.T. Both procedures lead to a multi-attribute utility function, but the viewpoints of risk attitude and methods of assessment are quite different.

The Stanford group views risk attitude as a one-dimensional phenomenom separate from deterministic tradeoffs among attributes. Figure 2.1 illustrates this decomposition approach. The deterministic preference axioms guarantee the existence of a deterministic preference function $V(\underline{x})$ that rank orders preferences over all possible multi-attribute outcomes. This preference function is mapped into a scalar order-preserving numeraire function $n(\underline{x})$. Using the risk preference axioms, the analyst assesses risk attitude over the one-dimensional numeraire to specify the multi-attribute risk preference function $u[n(\underline{x})]$.

Barrager and Keelin have written the most recent dissertations [3],[20] on multi-attribute utility theory in the Stanford decision analysis research program. Using the normative axioms (2.1-2.9) as a foundation, they specify additional assumptions that limit the form of the preference function. Barrager suggests properties over multi-period consumption preferences from which he derives the sum-of-exponentials, Cobb-Douglass, and sum-of-powers preference functions. Keelin generalizes Barrager's procedure and adds new preference parameters to encode utility functions for general multi-attribute problems. The additional assumptions they require for the sum-of-exponentials, sum-of-powers, and Cobb-Douglass functions are listed in Appendix B. These assumptions are restrictive since the same form of the preference parameter (the marginal value reduction coefficient) must hold everywhere. Of all the assumptions they propose that lead to specific families of curves, these three are the least

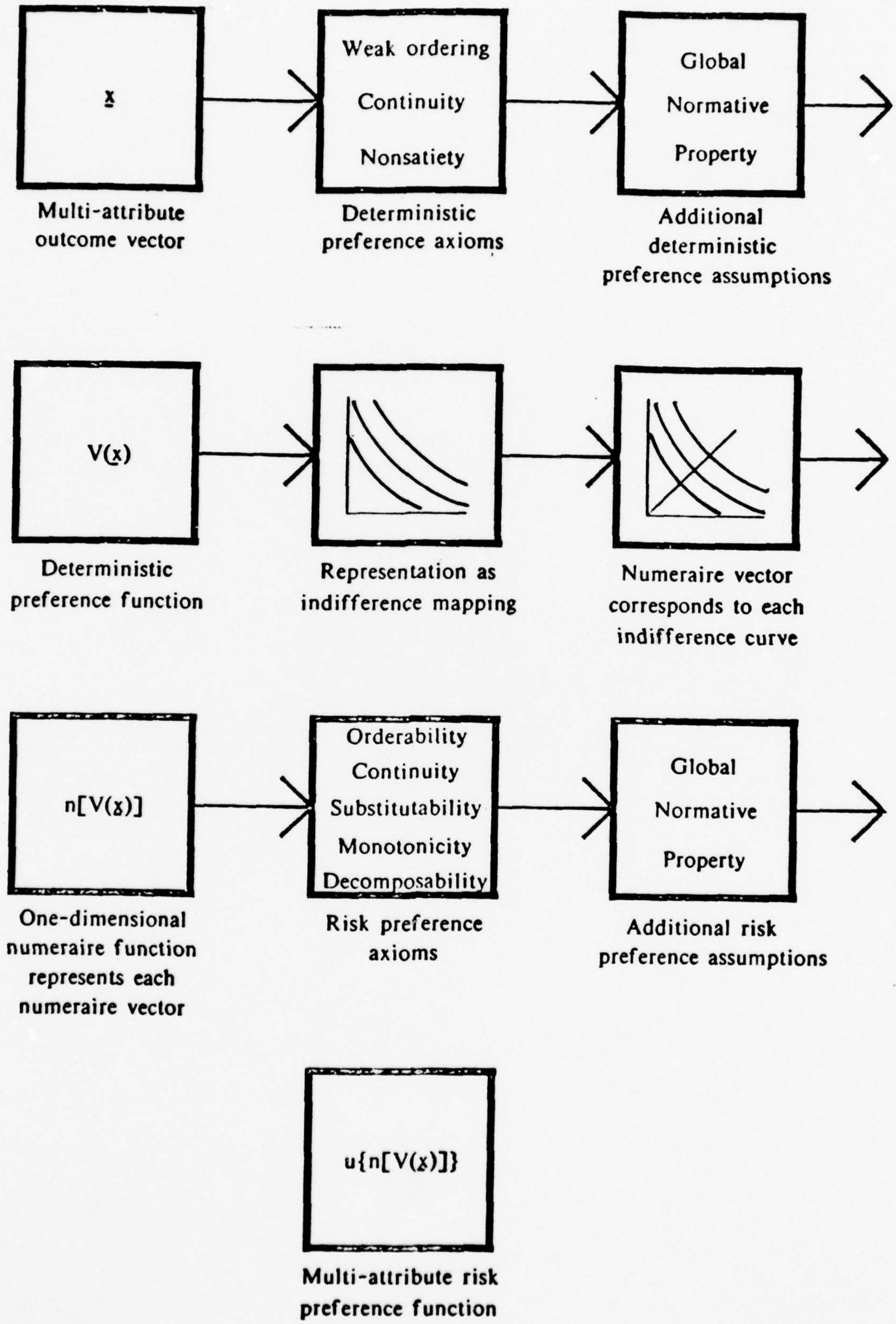


Figure 2.1 Decomposition of Deterministic and Risk Preference

restrictive. This global modeling technique provides mathematically simple functions, but it is built upon the restrictive assumptions we are trying to avoid.

Howard Raiffa and Ralph Keeney [21], at Harvard and M.I.T., and Peter Fishburn [11], at the Research Analysis Corporation, advocate a methodology in which deterministic and risk preferences are jointly assessed. They use a hierarchy of independence assumptions among subsets of attributes to restrict the preference function to additive or multiplicative forms. From the Stanford viewpoint, their procedure has one major drawback: it requires inference of deterministic tradeoffs from probabilistic questions. The risk additive independence and utility independence conditions from which they derive the additive and multiplicative utility functions are listed in Appendix B. My discussion of the Keeney-Raiffa methodology is brief since my research does not draw upon it directly, but a detailed account is available in their forthcoming book [21].

Conjoint measurement is a third approach leading to global deterministic preference functions. With this technique, the decision maker rank orders different combinations of attributes "considered jointly". A regression routine then assigns to each discrete level of each attribute a numerical value that minimizes the errors of the rank orderings according to an additive or multiplicative value model. The global restriction enters this technique through the underlying value model. Conjoint measurement and the closely related technique of multidimensional scaling have been applied recently to study consumer preferences in marketing research [19],[22],[24].

All three multi-attribute procedures described above are built upon strong assumptions that limit the form of the preference function. In the next section, I examine techniques designed to avoid these restrictions.

2.3 Boyd's Successive Approximation Algorithm

Dean Boyd [4] developed a multi-attribute procedure that does not construct a global preference function. Instead, it uses local models to generate a sequence of trial solutions that eventually converge to the optimum. We will first consider decision making under certainty to see how his procedure operates.

Boyd makes the following assumptions:

i. The deterministic preference axioms hold, so a concave and differentiable deterministic preference function $V(\underline{x})$ exists. However, $V(\underline{x})$ is unknown and assessment of the entire $V(\underline{x})$ is impossible.

ii. The attributes are modeled so negative quantities are meaningless; therefore $\underline{x} > \underline{0}$.

iii. The decision maker's resources are bounded by the convex constraint set $X = \{\underline{x} \mid \underline{h}(\underline{x}) = \underline{0},\ \underline{g}(\underline{x}) \leq \underline{0},\ \underline{x} \geq \underline{0}\}$. (This characterization of the feasible region is a standard nonlinear programming formulation that will be useful in establishing later results). The set of feasible decisions D is a subset of N-dimensional Euclidean space, $D = \{\underline{d} \mid \underline{x}(\underline{d}) \in X\}$.

Under these assumptions, Boyd tries to solve the following problem:

$$\begin{aligned} &\text{Maximize}_{\underline{d}} \quad V[\underline{x}(\underline{d})] \\ &\text{subject to} \quad \underline{d} \in D. \end{aligned} \tag{2.1}$$

This notation indicates he is trying to choose the decision $\underline{d}^*$ that leads to the most preferred outcome $\underline{x}(\underline{d}^*)$. Figure 2.2 illustrates the following simple two-dimensional case ($\underline{x}$ is understood to be a function of the decision $\underline{d}$):

$$\begin{aligned} &\text{Maximize}_{\underline{x}} \quad V(x_1,x_2) \\ &\text{subject to} \quad c_1x_1 + c_2x_2 \leq b \quad \text{and} \quad x_1,\ x_2 \geq 0. \end{aligned}$$

Pretending $V(\underline{x})$ is known, we draw its isovalue curves, projected onto the x_1x_2 plane. Each isovalue curve is a locus of points among which the decision maker is indifferent; economists refer to these loci as indifference curves. By the definition of $V(\underline{x})$, prospects on a higher curve, $V(\underline{x}) = k_2$, are preferred to prospects on a lower curve, $V(\underline{x}) = k_1$, where $k_2 > k_1$. With Axiom 2.4, we assumed the marginal rates of substitution are decreasing along each indifference curve. Since the slope $dx_1/dx_j\,|_{dV=0}$ is increasing in x_j, but always negative and finite, the indifference curves are convex to the origin. In this problem, Boyd is trying to maximize $V(\underline{x})$ subject to the constraints. He can find this maximum by identifying the point where an indifference

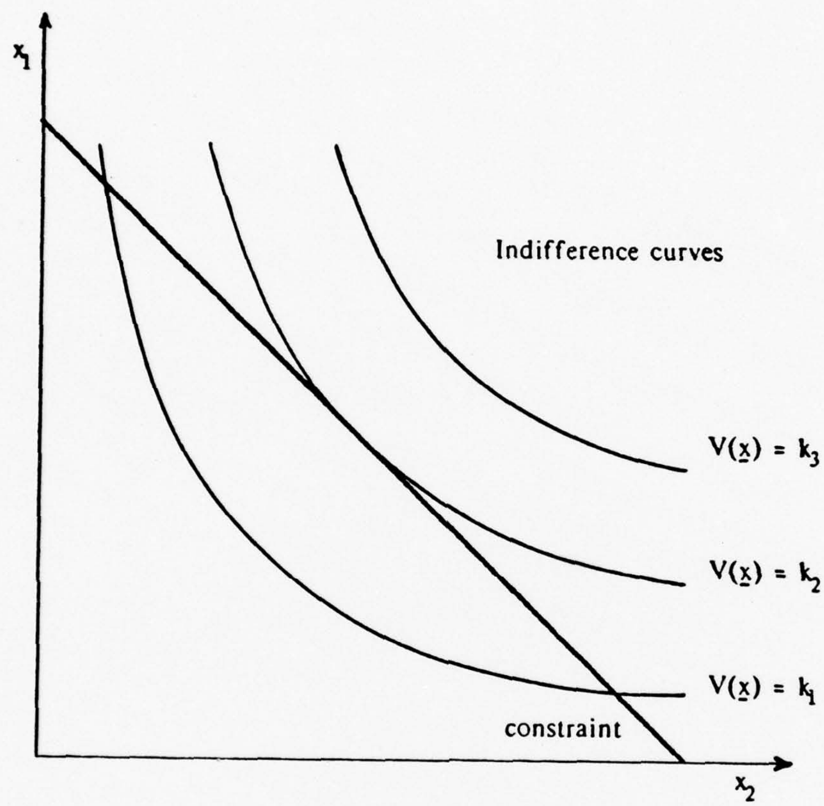


Figure 2.2 Graphical Illustration of Decision Problem

curve is just tangent to the constraint.

Boyd's first step is tradeoff assessment; he determines the Δx_1 at which the decision maker is indifferent between (x_1, x_j) and $(x_1-\Delta x_1, x_j+\Delta x_j)$. Figure 2.3 shows that this tradeoff $\Delta x_1/\Delta x_j$ is the negative slope of the tangent to the indifference curve. It expresses the amount of x_1 the decision maker is willing to sacrifice in order to gain a unit increment of x_j, leaving all other attributes constant. The tradeoff $\lambda_j[\underline{x}(\underline{d})]$ is the marginal rate of substitution of x_1 for x_j at $\underline{x}(\underline{d})$, using x_1 as a price variable:

$$\lambda_j(\underline{x}) = dx_1/dx_j \mid_{dV=0,\ dx_r=0,\ r\neq 1j}$$

To assess tradeoffs at any point $\underline{x}(\underline{d})$, the analyst presents the following prospects to the decision maker:

$$\underline{x} = [x_1,\ x_2,\ \ldots,\ x_j,\ \ldots,\ x_N]$$

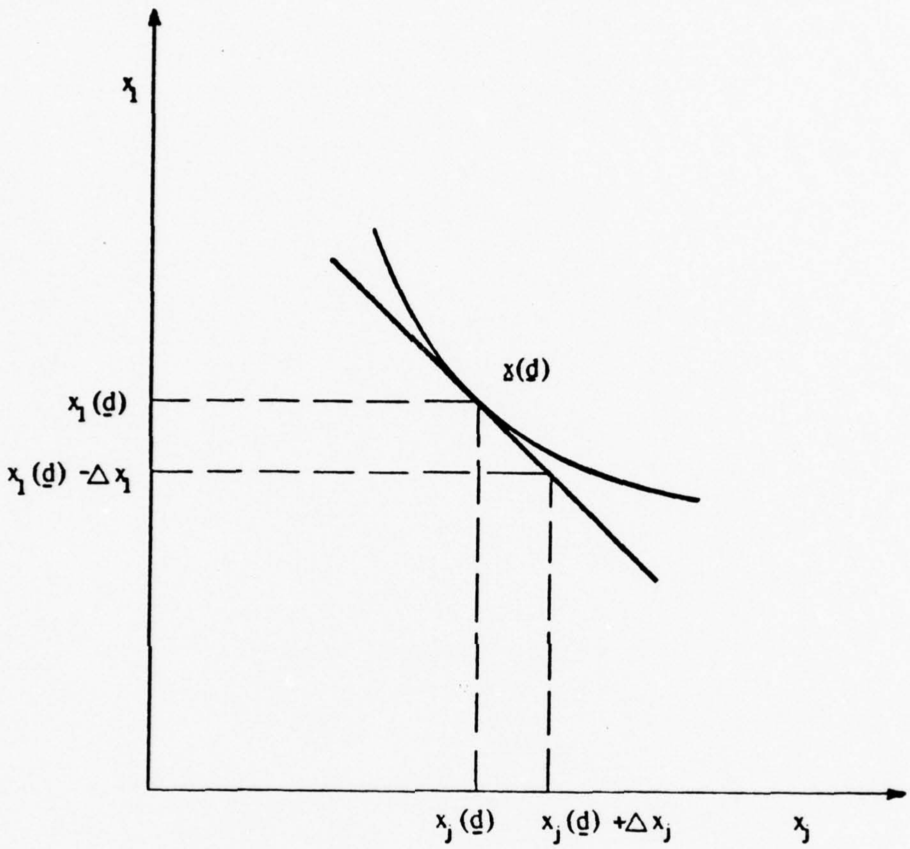
$$\underline{x}' = [x_1 - \Delta x_1,\ x_2,\ \ldots,\ x_j + \Delta x_j,\ \ldots,\ x_N]$$

For a small fixed Δx_j, *small enough so the indifference curve is approximately linear but large enough so the increment is meaningful*, the analyst varies Δx_1 until the decision maker is indifferent between $\underline{x}$ and $\underline{x}'$. At this level, $\lambda_j[\underline{x}(\underline{d})] \approx \Delta x_1/\Delta x_j$. Since λ_1 is always one, there are only N-1 degrees of freedom among tradeoffs at any point. Consistency can be checked by assessing a second set of tradeoffs with a different price variable x_k since the chain rule implies $\lambda_{1j}\,\lambda_{jk} \equiv \lambda_{1k}$.

Boyd defines the pseudo-objective function $h[\underline{x}(\underline{d})|\underline{x}(\underline{d}^k)]$ as a scaled linear approximation of the true preference function fit at $\underline{x}(\underline{d}^k)$. Taking the first-order Taylor expansion of V at $\underline{x}(\underline{d}^k)$, dividing by $[\partial V(\underline{x})/\partial x_1]_{\underline{x}=\underline{x}(\underline{d}^k)}$, and subtracting the constant terms, Boyd defines

$$\begin{aligned} h[\underline{x}(\underline{d})|\underline{x}(\underline{d}^k)] &= \textstyle\sum_i \left([\partial V(\underline{x})/\partial x_i]/[\partial V(\underline{x})/\partial x_1]|_{\underline{x}=\underline{x}(\underline{d}^k)}\right) x_i(\underline{d}) \\ &= \textstyle\sum_i \lambda_i[\underline{x}(\underline{d}^k)]\, x_i(\underline{d}) \end{aligned}$$

where λ_i is the tradeoff using x_1 as the price variable. This pseudo-objective is not the true preference function; it is a linear approximation valid in a small neighborhood of $\underline{x}(\underline{d}^k)$. Boyd proves the following theorem:



$$\lambda_j[\underline{x}(\underline{d})] \approx \Delta x_1 / \Delta x_j$$

Figure 2.3 The Marginal Rate of Substitution

Theorem 2.1. If the decision maker's preference ordering satisfies the deterministic preference axioms (2.1-2.4), and if $X(D)$ is convex, then if $\underline{d}^*$ maximizes $h[\underline{x}(\underline{d})|\underline{x}(\underline{d}^*)]$ over all $\underline{d} \in D$, then $\underline{d}^*$ also maximizes $V[\underline{x}(\underline{d})]$ for all $\underline{d} \in D$.

If the pseudo-objective fit at $\underline{d}^*$ achieves its maximum at $\underline{d}^*$, then the true objective also achieves its maximum at $\underline{d}^*$. Figure 2.4 shows that at a non-optimal decision $\underline{d}^n$, the gradient to the pseudo-objective has a component in the feasible region. However, at $\underline{x}(\underline{d}^*)$, there is no feasible direction of improvement, so $\underline{d}^*$ is the optimum. This simple illustration of the Kuhn-Tucker conditions (see Appendix C) shows maximization of the first order pseudo-objective at $\underline{d}^*$ is a sufficient condition for maximization of the true objective. Since the linear pseudo-objective is concave and the feasible region is convex, the condition is necessary as well. A formal proof of a more powerful version of this theorem is included in Chapter III.

This theorem by itself does not help solve the decision problem since it requires prior knowledge of the optimum. However, it serves as the backbone of Boyd's successive approximation algorithm outlined below:

Step 0. Choose an arbitrary $\underline{d}^0$ and assess tradeoffs. Let $k = 0$.

Step 1. Fit the pseudo-objective function using tradeoffs at $\underline{d}^k$ and maximize $h[\underline{x}(\underline{d})|\underline{x}(\underline{d}^k)]$ over all $\underline{d} \in D$. This maximization yields a new $\underline{d}^{k+1}$.

Step 2. If $\underline{d}^{k+1} = \underline{d}^k$, stop; $\underline{d}^{k+1}$ is the optimum (by Theorem 2.1).
If $\underline{d}^{k+1} \neq \underline{d}^k$, assess tradeoffs at $\underline{d}^{k+1}$. Let $k = k+1$ and return to Step 1.

This algorithm generates a sequence of points hopefully converging to the optimum. Figure 2.5 shows a simple example: the underlying indifference curves are shown for illustrative purposes (in our real problem, these curves are unknown). After arbitrarily selecting d^0 as an initial feasible point, the analyst assesses the marginal rate of substitution of x_1 for x_2 at $\underline{x}(d^0)$ to determine the local linear approximation $h[\underline{x}(d)|\underline{x}(d^0)]$. Figure 2.5a shows this linear pseudo-objective is maximized at $\underline{x}(d^1)$. Since $d^1 \neq d^0$, the new point $\underline{x}(d^1)$ is used in the next iteration. The tradeoff assessment at $\underline{x}(d^1)$ yields a new pseudo-objective $h[\underline{x}(d)|\underline{x}(d^1)]$ which achieves its maximum at $\underline{x}(d^2)$. Figure 2.5b shows $\underline{x}^2$ lies on a lower indifference curve, implying

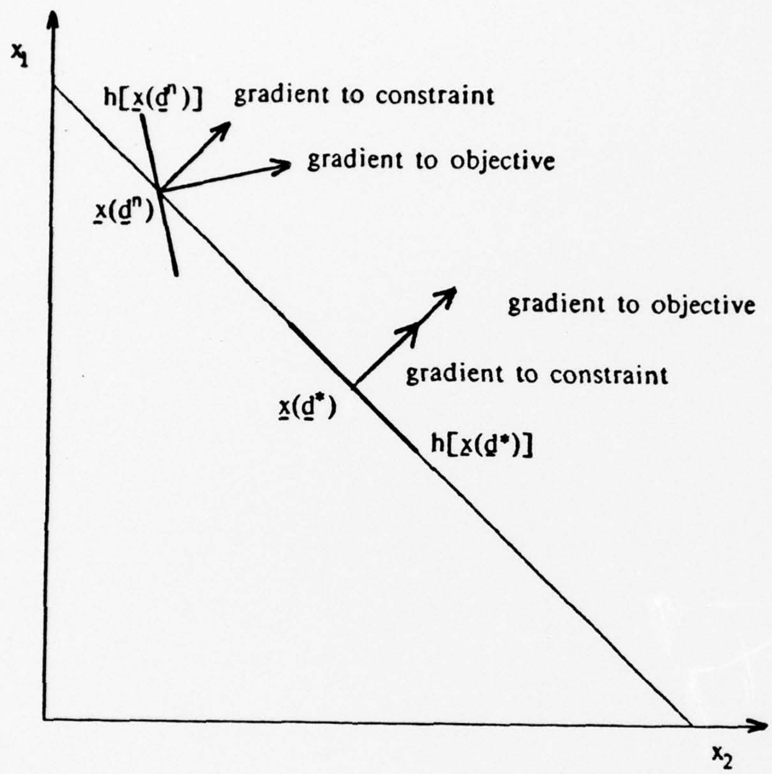
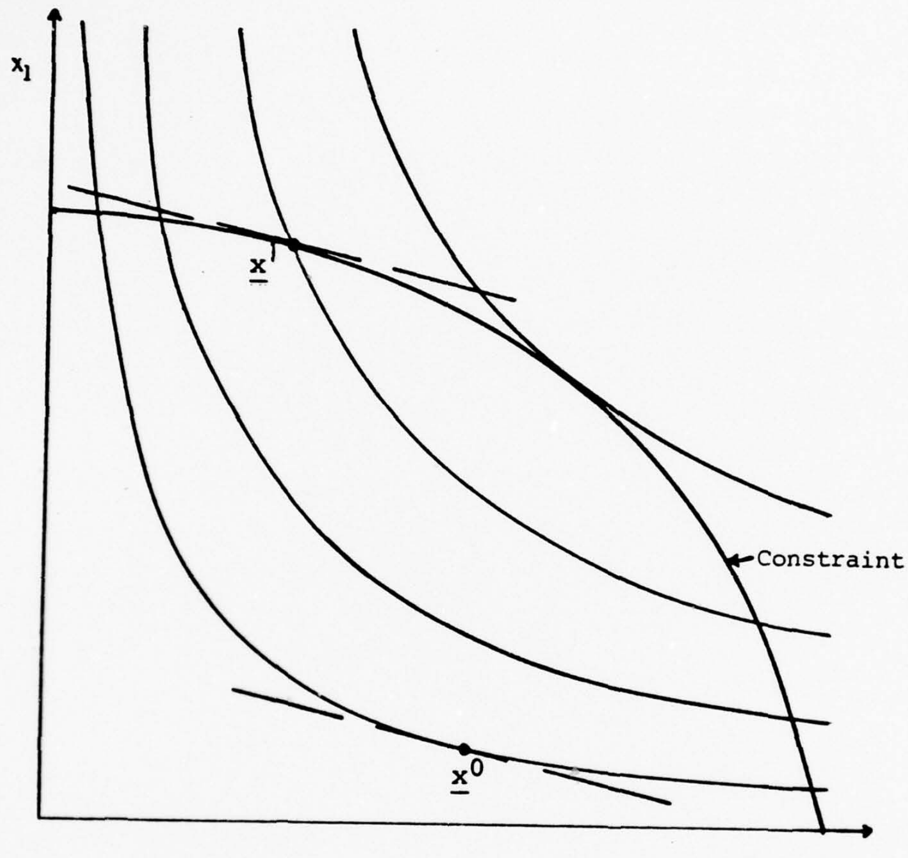


Figure 2.4 Two-dimensional Example of Theorem 2.1



First Iteration

Figure 2.5a Simple Example of Boyd's Algorithm

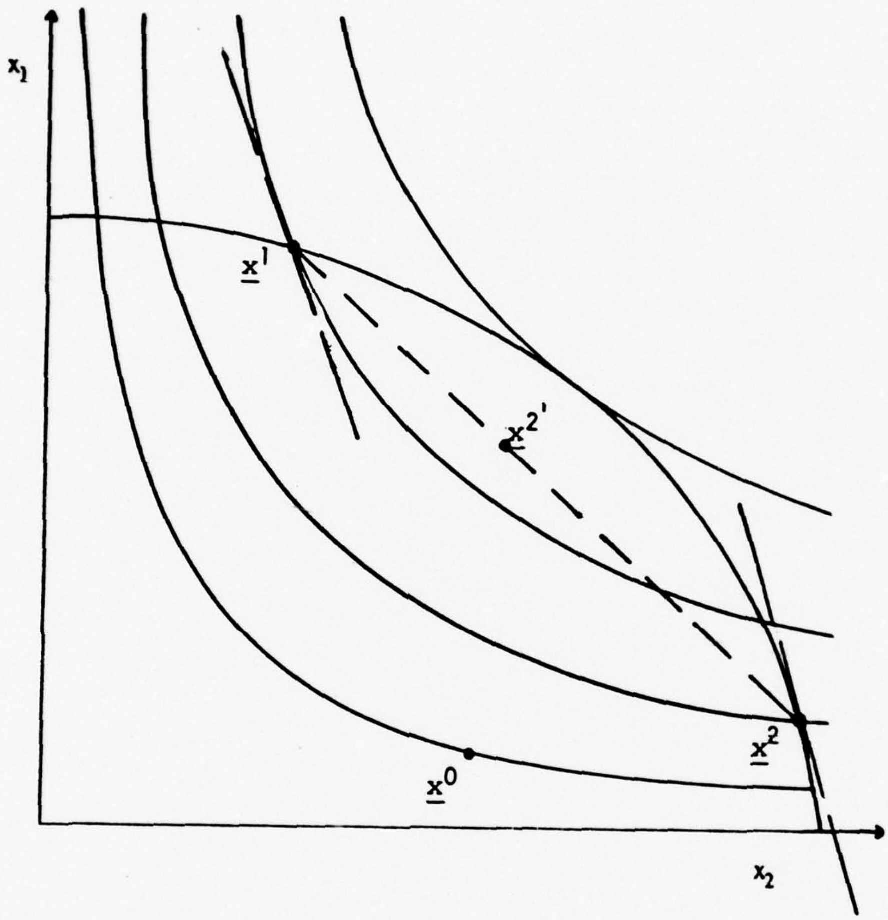


Figure 2.5b Simple Example: Second Iteration

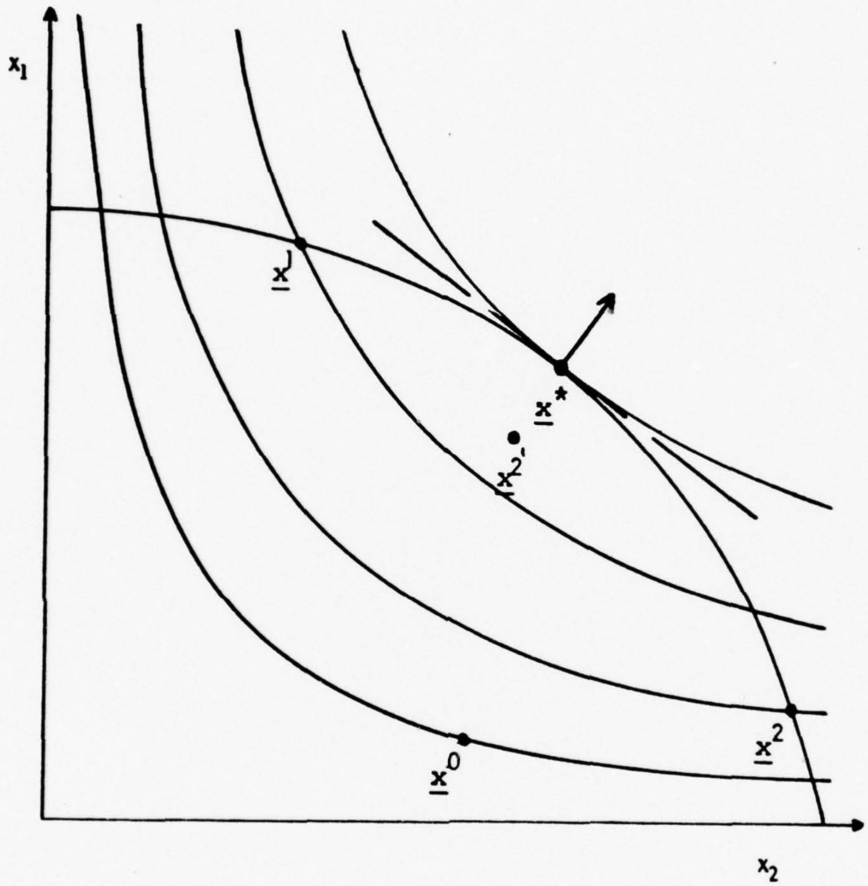


Figure 2.5c Simple Example: Final Iteration

$\underline{x}^1 \succ \underline{x}^2$. For the algorithm to converge, each iteration must provide a new point preferred to its predecessor. The prospect $\underline{x}^2$ is not an improvement, but it indicates a *feasible direction of improvement.* Consequently, a new point preferred to $\underline{x}^1$ can be chosen somewhere along the line segment from $\underline{x}^1$ to $\underline{x}^2$. Figure 2.5b shows a new point $\underline{x}^{2'}$ chosen in this manner; this new trial solution lies on a higher indifference curve than $\underline{x}^1$. The iterative procedure continues until $d^{k+1} = d^k$. Figure 2.5c shows that $h[\underline{x}(d)|\underline{x}(d^*)]$ is maximized at $\underline{x}(d^*)$; no further improvement can be made, so d^* is the optimum.

Successive linear approximation of a nonlinear objective function is not a new technique. The method was originally proposed by Frank and Wolfe [12] in 1956 for problems with linear constraints. In the original Frank-Wolfe algorithm, each iteration used an exact line search to find the next point. The procedure has since been applied with inexact line searches to problems with linear and nonlinear constraints. Boyd's contribution was the application of this modified Frank-Wolfe algorithm to the multi-attribute decision in which the preference function is unknown. Unable to evaluate the gradient of V at each $\underline{x}^k$ (since V is unknown), he assesses the decision maker's tradeoffs at $\underline{x}^k$. These tradeoffs provide a vector $\underline{\lambda}(\underline{x}^k)$ collinear with $\nabla V(\underline{x}^k)$ since

$$\lambda_j(\underline{x}^k) \quad = \quad [\partial V(\underline{x}^k)/\partial x_j]/[\partial V(\underline{x}^k)/\partial x_1].$$

Maximizing the pseudo-objective $\underline{\lambda}(\underline{x}^k)^T\underline{x}$ is equivalent to maximizing $\nabla V(\underline{x}^k)\underline{x}$ in the Frank-Wolfe algorithm since the scaling factor $\partial V(\underline{x}^k)/\partial x_1$ is a positive constant.

In summary, Boyd's procedure avoids global preference modeling by using a sequence of tradeoff assessments, local linear models, and optimizations. However, Boyd pays a heavy price to avoid the restrictive assumptions; his iterative procedure requires many more assessments and optimizations. It converges very slowly in general since the linear pseudo-objectives are very poor preference models even in the small. The rate of convergence decreases with increasing nonlinearity of the true preference function; in many problems the true preference function is highly nonlinear. We must ask ourselves if Boyd's algorithm solves our basic problem. In theory, it does, but in practice, it does

not, since the large number of iterations makes the assessment demands prohibitive.

Geoffrion, Dyer, and Feinberg [8],[14] independently developed an interactive procedure similar to Boyd's successive approximation algorithm. Their scheme also uses the Frank-Wolfe linear approximation technique at each iteration. However, they do not ask the decision maker if $\underline{x}^{k+1}$ is preferred to $\underline{x}^k$. Rather, they require the decision maker to choose his most preferred point along the line segment from $\underline{x}^k$ to $\underline{x}^{k+1}$. The key difference is that Geoffrion, Dyer, and Feinberg's procedure requires the decision maker to choose among many outcomes all at once, in contrast to Boyd's scheme requiring comparison of only two points at a time. The decision maker's task is more difficult, but the relaxation procedure is avoided. Dyer [9] and Hogan [15] investigate the convergence of the modified Frank-Wolfe method using an approximation of the true gradient at each step. I will use several of their convergence analyses in Chapters III and IV.

Wehrung [33] establishes theoretical foundations for interactive identification and optimization of preferences. For both numerical and qualitative forms of preference, he examines elicitation procedures and mathematical programming techniques that guide the search and indicate trial points where the decision maker should identify his preferences. Tests monitoring the consistency of the assessed information with the basic rationality conditions are also developed. Wehrung's results are entirely theoretical; he makes no attempt in his dissertation to develop a practical procedure.

My dissertation draws bits and pieces from most of the literature reviewed in this chapter and from optimization theory, but the primary motivation was Dean Boyd's thesis [4]. I felt the same dissatisfaction with restrictive state-of-the-art techniques that he felt. I thought his alternative approach was clever and imaginative, and in this dissertation, I try to make it operational.

CHAPTER III

THE PROXY ITERATION ALGORITHM FOR DECISION MAKING UNDER CERTAINTY

3.1 New Proxy

The pseudo-objective governs the speed at which the successive approximation algorithm converges. At each iteration, the pseudo-objective is maximized as if it were the true objective and a new trial point is found. Since Boyd's linear functions are very poor pseudo-objectives, the trial points they generate are very poor pseudo-optima. As a result, the sequence progresses very slowly toward the true optimum.

I propose to use the preference functions derived by Barrager [3], and generalized by Keelin [20], as local proxies in the successive approximation algorithm. As a first step, I use the global assessment procedure to encode a preference function, but instead of using this function as a global model, I use it only as a local proxy. At each iteration, a new tradeoff vector is assessed to update the proxy. Since the preference function is a very good model in the small, the algorithm should converge at a much higher rate. In this proposed procedure, the proxy is never assumed to be the true preference function, even in the small; it serves only as a mechanism guiding the search for the optimal decision.

Even though we do not use Barrager's functions as global models, we still must be convinced that they are suitable local proxies. Barrager provides normative motivations for the sum-of-exponentials, sum-of-powers, and Cobb-Douglass preference functions in a time-preference context and Keelin restates them in more general terms. Appendix B includes a brief summary of their arguments.

To demonstrate the new algorithm, I will develop my methodology using the sum-of-exponentials preference function as a local proxy to solve problem (2.1). I call the local model a proxy rather than a pseudo-objective.

Definition 3.1. Let $p[\underline{x}(\underline{d})]$ be the sum-of-exponentials approximation of the deterministic preference function $V[\underline{x}(\underline{d})]$;

$$p[\underline{x}(\underline{d}) \mid \underline{x}(\underline{d}^n), \underline{x}(\underline{d}^{n-1})] = -\sum_i a_i e^{-\omega_i x_i(\underline{d})},$$

where a_i and ω_i are determined from tradeoff assessments at $\underline{x}(\underline{d}^n)$ and $\underline{x}(\underline{d}^{n-1})$ such that $\nabla p[\underline{x}(\underline{d}^n)]$ is collinear with $\nabla V[\underline{x}(\underline{d}^n)]$, and $\nabla p[\underline{x}(\underline{d}^{n-1})]$ is collinear with $\nabla V[\underline{x}(\underline{d}^{n-1})]$.

Theorem 3.1. If the decision maker's preference ordering satisfies the deterministic preference axioms (2.1-2.4), if $X(D)$ is convex, and if $\underline{x}(\underline{d}^*)$ is a regular point of the constraints (see Appendix B), then if $\underline{d}^*$ maximizes $p[\underline{x}(\underline{d}) \mid \underline{x}(\underline{d}^*), \underline{x}(\underline{d}^k)]$ over all $\underline{d} \in D$, for any $\underline{x}(\underline{d}^k)$, then $\underline{d}^*$ also maximizes $V[\underline{x}(\underline{d})]$ for all $\underline{d} \in D$.

If the proxy fit at $\underline{d}^*$ achieves its maximum at $\underline{d}^*$, then the true objective also achieves its maximum at $\underline{d}^*$. This theorem generalizes Theorem 2.1 since it holds for any concave proxy, the linear proxy being the simplest case.

Figure 3.1 illustrates the idea of Theorem 3.1 in a two-dimensional example. If the gradient to the objective has a component in the feasible region, further improvement can be made, so $\underline{x}(\underline{d}^n)$ is not optimal. If the gradient to the objective is perpendicular to the constraint, as at $\underline{x}(\underline{d}^*)$, no direction of improvement is feasible, so $\underline{d}^*$ is optimal. Figure 3.1 gives a simple interpretation of the Kuhn-Tucker optimality conditions (see Appendix C). The preference function $V(\underline{x})$ is never specified; only vectors collinear with its gradient at a few points are known. When maximizing the concave objective with the convex constraint set, this first-order information is both necessary and sufficient to guarantee optimality. Nonconvex feasible regions are discussed in section 3.4.

Proof: The proxy $p(\underline{x})$ and the true objective $V(\underline{x})$ are concave, the constraint set $X(D) = \{\underline{x} \mid \underline{h}(\underline{x}) = \underline{0},\ \underline{g}(\underline{x}) \leq \underline{0},\ \underline{x} \geq \underline{0}\}$ is convex, and $\underline{x}(\underline{d}^*)$ is a regular point of the constraints. Therefore, if $\underline{d}^*$ maximizes $p[\underline{x}(\underline{d}) \mid \underline{x}(\underline{d}^*), \underline{x}(\underline{d}^k)]$ for all $\underline{d} \in D$, for any $\underline{x}(\underline{d}^k)$, the Kuhn-Tucker necessary conditions guarantee the existence of $\underline{\lambda}$ and $\underline{\mu}$, $\underline{\mu} \geq \underline{0}$, such that

$$\nabla p(\underline{x}^*) + \underline{\lambda} \nabla \underline{h}(\underline{x}^*) + \underline{\mu} \nabla \underline{g}(\underline{x}^*) = \underline{0}.$$

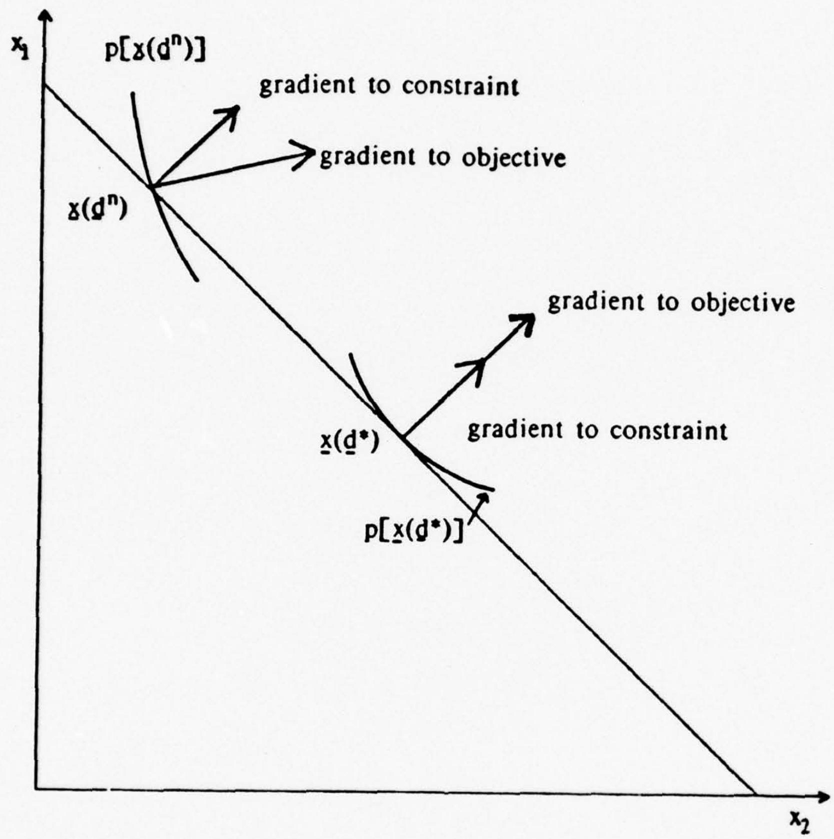


Figure 3.1 Two-dimensional Example of Theorem 3.1

By construction,

$$\nabla p(\underline{x}^*) = t\nabla V(\underline{x}^*)$$

for some positive scalar t; by defining $\underline{\tau} = (1/t)\underline{\lambda}$ and $\underline{\nu} = (1/t)\underline{\mu}$, we have

$$\nabla V(\underline{x}^*) + \underline{\tau}\nabla \underline{h}(\underline{x}^*) + \underline{\nu}\nabla g(\underline{x}^*) = \underline{0}, \quad \underline{\nu} \geq \underline{0}.$$

But this equation satisfies the Second-Order Sufficiency Conditions (Appendix C) since V is concave and $X(D)$ is convex, so $\underline{x}(\underline{d}^*)$ maximizes the true objective. These identical steps, applied in reverse, prove the converse. Q.E.D.

Theorem 3.1 holds for any concave approximation that is fit from a vector collinear with $\nabla V(\underline{x})$ at $\underline{x}(\underline{d}^*)$ and that satisfies the deterministic preference axioms. This generalization is true since the optimality criterion requires only the gradient of the proxy at $\underline{x}(\underline{d}^*)$. Therefore, the sum-of-powers preference function, $V(\underline{x}) = \Sigma_i a_i x_i^{-\alpha_i}$, and the Cobb-Douglass preference function, $V(\underline{x}) = \Sigma_i a_i \ln x_i$, could also serve as proxies for $V(\underline{x})$.

Just as in Boyd's development, this theorem by itself does not solve the decision problem since it requires prior knowledge of the optimum. However, it motivates a new successive approximation algorithm, using the sum-of-exponentials proxy at each iteration:

Step 0. Choose an arbitrary $\underline{d}^0$ and $\underline{d}^1$ and assess tradeoffs. Let $k = 1$.

Step 1. Fit a sum-of-exponentials proxy using tradeoffs at $\underline{d}^k$ and $\underline{d}^{k-1}$ and maximize $p[\underline{x}(\underline{d}) \mid \underline{x}(\underline{d}^k), \underline{x}(\underline{d}^{k-1})]$ over all $\underline{d} \in D$. This maximization yields a new $\underline{d}^{k+1}$.

Step 2. If $\underline{d}^{k+1} = \underline{d}^k$, stop; $\underline{d}^{k+1}$ is the optimum (by Theorem 3.1).
If $\underline{d}^{k+1} \neq \underline{d}^k$, assess tradeoffs at $\underline{d}^{k+1}$. Let $k = k + 1$ and return to Step 1.

I call this new procedure the *proxy iteration algorithm*. The algorithm as written here is not guaranteed to converge. However, it is fail-safe since if it does converge, the result is the true optimum. Special devices will be added in section 3.4 to guarantee global convergence. In order to develop the algorithm in its simplest form, we assume

initially, in sections 3.2 through 3.4, that the decision maker's tradeoffs are consistent with a deterministic preference function satisfying Axioms 2.1-2.4. We then relax this assumption in Section 3.5 and examine the effects of assessment error.

3.2 Information Requirements of the New Proxy

The sum-of-exponentials function is a higher-order model and requires more parameters to fit. Since an ordinal preference function is unique up to an increasing monotonic transformation, the constant a_1 can arbitrarily be set equal to one in $p(\underline{x}) = -\Sigma_i a_i e^{-\omega_i x_i}$. The remaining 2N-1 parameters, $a_2, \ldots, a_N, \omega_1, \omega_2, \ldots, \omega_N$, must be calculated from tradeoff assessments. At any $\underline{x}$, there are N-1 tradeoffs $\lambda_i(\underline{x})$, $i = 2, \ldots, N$, since $\lambda_1 = 1$ when x_1 is the price variable. A full set of N-1 tradeoffs at each of two points plus a single tradeoff at a third point are required to fit the 2N-1 parameters. The numerical tradeoffs λ_i actually assessed relate to the sum-of-exponentials parameters a_i and ω_i in the following way:

$$\lambda_i(\underline{x}) = [\partial p(\underline{x})/\partial x_i] \;/\; [\partial p(\underline{x})/\partial x_1] = (\omega_i a_i e^{-\omega_i x_i}) \;/\; (\omega_1 a_1 e^{-\omega_1 x_1}),$$

so the ratio of $\lambda_i(\underline{x}^1)$ to $\lambda_i(\underline{x}^2)$ is

$$[\lambda_i(\underline{x}^1)]/[\lambda_i(\underline{x}^2)] = e^{\{\omega_1(x_1^1 - x_1^2) - \omega_i(x_i^1 - x_i^2)\}}.$$

Taking the logarithm of both sides,

$$\ln [\lambda_i(\underline{x}^1)/\lambda_i(\underline{x}^2)] = \omega_1(x_1^1 - x_1^2) - \omega_i(x_i^1 - x_i^2), \quad i = 2, 3, \ldots, N.$$

Since there are N-1 equations and N unknowns, one more $\lambda_j(\underline{x}^3)$ is assessed to provide a second equation in ω_1 and ω_j,

$$\ln [\lambda_j(\underline{x}^1)/\lambda_j(\underline{x}^3)] = \omega_1(x_1^1 - x_1^3) - \omega_j(x_j^1 - x_j^3).$$

Solving for ω_1 by Cramer's rule,

$$\omega_1 = \Big[(x_j^3 - x_j^1) \ln [\lambda_j(\underline{x}^1)/\lambda_j(\underline{x}^2)] + (x_j^1 - x_j^2) \ln [\lambda_j(\underline{x}^1)/\lambda_j(\underline{x}^3)]\Big] \Big/ \Big[(x_1^1 - x_1^2)(x_j^3 - x_j^1) + (x_1^1 - x_1^3)(x_j^1 - x_j^2)\Big]$$

and

$$\omega_i = \left(\omega_1(x_1{}^1 - x_1{}^2) - \ln\,[\lambda_i(\underline{x}^1)/\lambda_i(\underline{x}^2)]\right) \,/\, (x_i{}^1 - x_i{}^2), \quad i = 2, 3, \ldots, N.$$

This $\underline{\omega}$ vector then determines the weighting factors $\underline{a}$,

$$a_i = [\lambda_i(\underline{x}^2)]\ (\omega_1/\omega_i)\ e^{\omega_i x_i{}^2 - \omega_1 x_1{}^2}, \quad i = 2, 3, \ldots, N, \quad a_1 = 1.$$

At first thought, doubling the information requirements of the proxy seems to add a considerable assessment burden. *However, the extra parameters are already available since tradeoffs are assessed at each iteration. Instead of throwing away the past information as in Boyd's procedure, my method uses tradeoffs at the current and previous points.* Therefore, the sum-of-exponentials proxy has no additional information cost.

This choice of proxy function illustrates an important optimization concept. In any search procedure, we want to use the most efficient method; we must weigh our desire to use the best model against its information requirements. The model selection itself is an optimization problem. In our multi-attribute decision, normative assumptions motivate the sum-of-exponentials proxy. It is a much better model than Boyd's linear approximation, but requires twice as much information. The situation here is rather unique, however, since the extra information is provided at no extra cost. The algorithm generates all the required information, so we have no delicate balance to consider. *The higher-order model is superior since it makes more efficient use of the information already available.*

An analogy can be drawn to the minimization of a polynomial using its Taylor series expansion. The method of steepest descent requires just the gradient, but achieves only first-order convergence. Newton's method requires the entire Hessian matrix, but achieves at least second-order convergence. The proxy iteration algorithm therefore plays the same role relative to the Frank-Wolfe procedure that Newton's method plays relative to steepest descent.

All these search procedures use models of the true objective at each stage. In our decision problem, we use local proxies to avoid the restrictions of global preference modeling. In other optimization problems, we use Taylor expansion models to avoid the

more difficult direct optimization of the true objective. In any case, if the model were identical to the true objective, the algorithm would converge in one step. Since the model is only a proxy for the true objective, each optimization yields only a trial solution for the true optimum; the better the proxy function fits the true objective, the faster the trial sequence reaches the optimum.

3.3 Optimizing the New Proxy

Nonlinear optimization problems are generally more difficult than linear optimization problems. Since we replace Boyd's linear pseudo-objective with the nonlinear proxy, we would expect each maximization to be more complicated. In the case of linear constraints, each iteration would require a nonlinear rather than a linear program.

The sum-of-exponentials, sum-of-powers, and Cobb-Douglass preference functions, however, all have a special mathematical structure that simplifies the task. All are *concave* and *separable*, sc the maximization with linear constraints, when converted to a minimization in standard form, becomes a convex separable problem. For the sum-of-exponentials proxy, the problem is written below:

$$\text{Maximize}_{\underline{x}} \quad -\sum_i a_i e^{-\omega_i x_i} \quad \ni \quad A\underline{x} \leq \underline{b}$$

or equivalently,

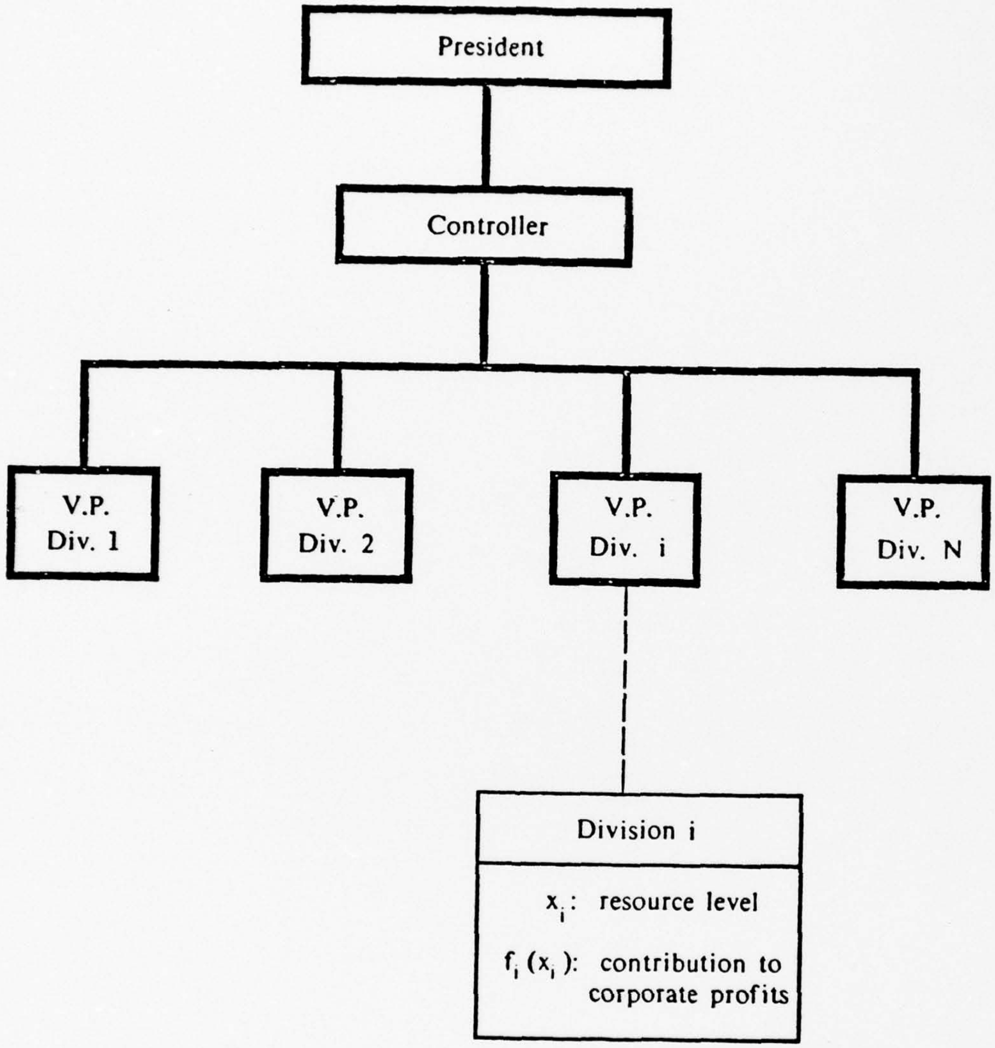
$$\text{Minimize}_{\underline{x}} \quad \sum_i a_i e^{-\omega_i x_i} \quad \ni \quad A\underline{x} \leq \underline{b}.$$

Convex separable techniques, using a series of linear programs, make this problem relatively easy to solve. Most multi-attribute decisions can be modeled effectively with a few attributes; a large number of attributes would make a problem unmanageable. For problems of small dimension, the new proxy with the convex programming algorithm so drastically decreases the number of iterations that even though each iteration takes a little longer, the total computer time is considerably reduced. This special mathematical structure and the algorithms designed to exploit it eliminate any computational burden associated with the concave proxy.

The multi-attribute decision with one linear budget constraint is a frequently occurring problem. Dual methods provide a closed-form solution to this special case. Before using the primal-dual technique [25],[26], I motivate the duality concept with the intuitive example in Figure 3.2.

The drawing represents the top management structure of a corporation. For the i^{th} division, x_i is the quantity of resource used and $f_i(x_i)$ is the profit from operating at level x_i. The president wants to maximize total corporate profits $\Sigma_i f_i(x_i)$ subject to the budget constraint $\Sigma_i x_i \leq b$. He is trying to allocate resources as efficiently as possible. He may find the optimal $(x_1^*, x_2^*, \ldots, x_N^*)$ himself and give orders to his vice-presidents; this centralized style of management corresponds to primal methods of optimization. In this problem, primal methods lead to $N+1$ simultaneous nonlinear equations in $N+1$ variables, a very difficult system to solve. Instead of solving the problem himself, the president may decide to let his controller set a shadow price μ for the resource. The vice-presidents would then choose their divisional operating levels and would be charged μ dollars for each unit of resource they use. Since the corporate profit function is concave separable, each vice president could independently choose the x_i that maximizes his division's profit $f_i(x_i) - \mu x_i$.

In this decentralized scheme, the controller is allocating corporate earnings to the income statements of the individual divisions. He tries to minimize the divisions' credits, making the earning power appear to be a corporate rather than divisional phenomenom. He allocates total credits μb by maximizing his share, $\mu b - \Sigma_i E_i(\mu)$, with respect to μ, where $E_i(\mu) = f_i[x_i^*(\mu)] - \mu x_i^*(\mu)$, the earnings of division i parametrized in terms of the shadow price μ. This technique requires $N+1$ one-dimensional suboptimizations, each of which can be solved by elementary calculus. The Convex Duality Theorem (see Appendix C) guarantees that dual methods, using this competitive situation between the controller and the vice-presidents, give the same solution the president would have reached using primal methods directly. The advantage of the dual technique is its computational simplicity. The closed-form solution for the sum-of-exponentials function with one linear budget constraint is derived as follows:



The President's Problem: Maximize $\Sigma f_i(x_i)$ subject to $\Sigma x_i \leq b$

Figure 3.2 Decentralization: An Intuitive View of Duality

$$\text{Minimize}_{\underline{x}} \quad f(\underline{x}) = \sum_i a_i e^{-\omega_i x_i}$$

$$\text{subject to} \quad \sum_i c_i x_i \leq b, \quad \text{where } b > 0 \quad \text{and} \quad a_i, \omega_i, c_i > 0 \ \forall i.$$

For all $\underline{x} \in X$, the Hessian of the objective $F(\underline{x})$ is positive definite and the constraint Hessian $G(\underline{x}) = [0]$. Therefore, the Lagrangian $L(\underline{x})$ is positive definite everywhere so any local minimum is a global mimimum. Using the Convex Duality Theorem (Appendix C), the dual function is

$$\varphi(\mu) = \min_{\underline{x}} \left[\sum_i a_i e^{-\omega_i x_i} + \mu\left(\left(\sum_i c_i x_i\right) - b\right)\right].$$

Separability implies each x_i can be optimized independently,

$$\min_{x_i} \ (a_i e^{-\omega_i x_i} + \mu c_i x_i), \qquad \mu \geq 0.$$

Setting the first derivative equal to zero yields

$$x_i = (-1/\omega_i) \ln [(\mu c_i)/(\omega_i a_i)]. \tag{3.1}$$

Substituting x_i into $\varphi(\mu)$ and simplifying yields

$$\varphi(\mu) = \mu \left[\sum_i \left((c_i/\omega_i) + (c_i/\omega_i) \ln [(\omega_i a_i)/(\mu c_i)]\right) - b\right].$$

Maximizing $\varphi(\mu)$ unconstrained, $\varphi'(\mu) = 0$ at

$$\mu^* = e^{\left[\left(\sum_i (c_i/\omega_i) \ln [(\omega_i a_i)/c_i]\right) - b\right] / \left(\sum_i c_i/\omega_i\right)}.$$

Substituting μ^* in (3.1) yields the closed-form solution

$$x_i^* = \left[b + \left(\ln [(\omega_i a_i)/c_i] \sum_i c_i/\omega_i\right) - \left(\sum_i (c_i/\omega_i) \ln [(\omega_i a_i)/c_i]\right)\right] / \left[\omega_i\left(\sum_i c_i/\omega_i\right)\right]. \tag{3.2}$$

The nonnegativity requirements were suppressed in the derivation above, so the solution $\underline{x}^*$ must now be checked for violations. A technique illustrating the central theme of duality can be used to insure nonnegativity. From equation (3.1), we can derive the parametric shadow price

$$\mu = (\omega_i a_i e^{-\omega_i x_i}) / c_i.$$

This equation shows that μ is the ratio of the marginal revenue of each division to its marginal cost. Since $\underline{\omega}$, $\underline{a}$, $\underline{c} > \underline{0}$,

$$\mu = (\omega_i a_i)/c_i \rightleftarrows x_i = 0;$$

$$\mu < (\omega_i a_i)/c_i \rightleftarrows x_i > 0;$$

$$\mu > (\omega_i a_i)/c_i \rightleftarrows x_i < 0.$$

The convex dual technique finds the $\underline{x}^*$ at which the shadow price is equal for all divisions. Figure 3.3 shows, in the context of the corporate example, that marginal revenue $\omega_i a_i e^{-\omega_i x_i}$ equals marginal cost c_1 at $x_i > 0$; at this level of x_i, the shadow price μ_1 is less than $(\omega_i a_i)/c_1$. For marginal cost c_2, the marginal revenue curve intersects marginal cost c_2 at $x_i < 0$; here the shadow price μ_2 is greater than $(\omega_i a_i)/c_2$. If equation (3.2) prescribes any $x_j < 0$, hence $(\omega_j a_j)/c_j < \mu$, the j^{th} vice president would reset x_j to zero to eliminate his division's loss (all costs are variable costs). The entire allocation would then be solved again, with the added restriction $x_j = 0$.

To handle nonnegativity violations efficiently, we rank the quantities $(\omega_i a_i)/c_i$ in ascending order. All are positive since $\underline{\omega}$, $\underline{a}$, $\underline{c} > \underline{0}$. If any violations occur, the x_j corresponding to the lowest $(\omega_j a_j)/c_j$ is set equal to zero. The shadow price μ is then recalculated excluding this x_j. If the next solution is nonnegative, it is optimal; if not, the x_j corresponding to the next lowest $(\omega_j a_j)/c_j$ is set to zero and the routine is repeated. Once an x_j is eliminated, it can never be made positive in a subsequent allocation since its associated $(\omega_j a_j)/c_j$ will always be less than the subsequent μ's. In most practical problems, intelligent modeling of the attributes will prevent negative solutions, but this dual technique will resolve any violations should they occur.

Problems with two or more constraints do not have closed-form solutions even with dual methods since the Lagrange multipliers themselves are entangled in simultaneous equations. If the constraints are linear, these problems can be solved easily with convex programming routines. Nonlinear constraints, however, with either Boyd's linear pseudo-objective or my concave proxy, require more complex nonlinear methods.

Having solved the optimization at each step, we conclude that just as the

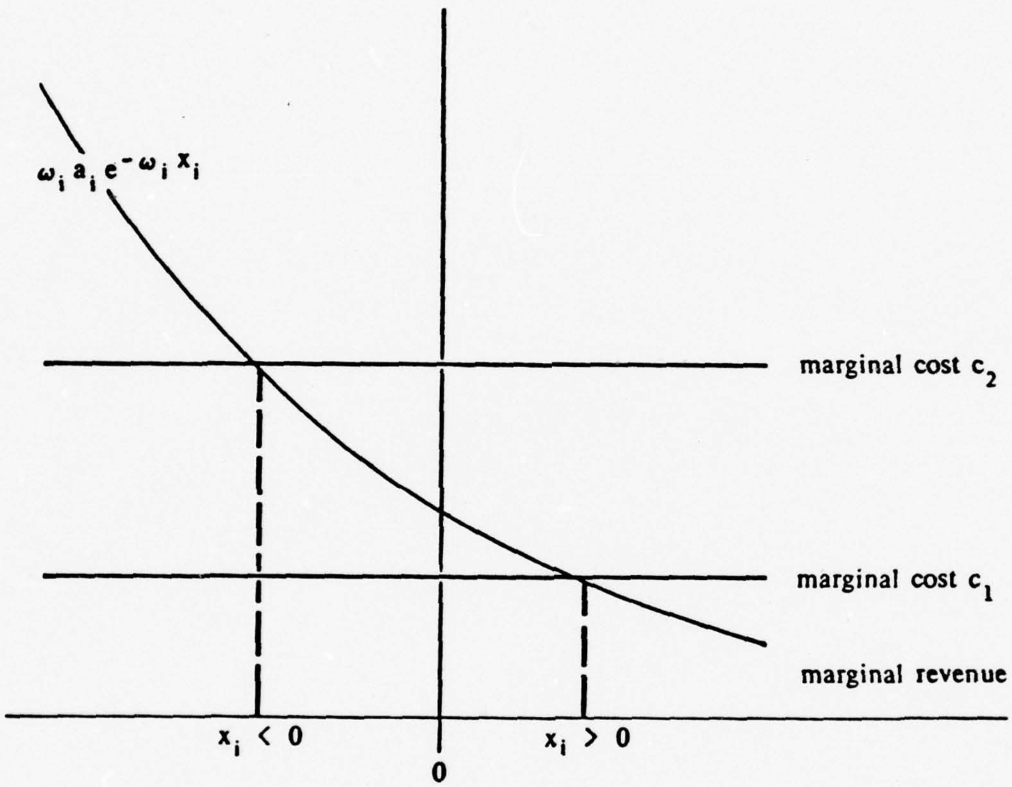


Figure 3.3 Duality: Marginal Revenue vs. Marginal Cost

higher-order proxy had no additional information cost, it adds no significant computational cost either. The maximization procedures for the Cobb-Douglass and sum-of-powers proxies are very similar since these functions are also concave separable. The special form of the Cobb-Douglass function allows a closed-form solution even with a quadratic constraint. The derivations with these functions follow the same steps, so they are not reproduced here, but the closed-form solutions will be used in Chapter IV.

3.4 Global Convergence with the New Proxy

The new proxy accelerates the successive approximation algorithm, but does not ensure convergence. To guarantee that the algorithm will always converge, we must verify that each iteration makes a sufficient improvement. Since we can only maximize the true objective by proxy, we must ask the decision maker after each maximization if the new point is preferred to the old one. The decision maker cannot specify the entire true objective, but we assume he can answer simple choice questions comparing two prospects. If he prefers the new point to the old, the next iteration may begin. If the new point is worse, something must be done to find a better one.

At the k^{th} iteration, the proxy is fit from vectors collinear with the gradients of the true objective at $\underline{x}^k$ and $\underline{x}^{k-1}$. The following theorem shows that at each iteration, the gradient of the true objective has a positive component along the direction of search obtained by maximizing the proxy. Optimizing the proxy, therefore, guarantees a search along a direction of genuine improvement.

Theorem 3.2. Given the decision problem (2.1), let $\underline{x}^{k+1} = \max_{\underline{x}} p[\underline{x} \mid \underline{x}^k, \underline{x}^{k-1}]$ for all $\underline{x} \in X(D)$. Then $\nabla V(\underline{x}^k)(\underline{x}^{k+1} - \underline{x}^k) > 0$.

Proof: At any $\underline{x}^k$ generated by the algorithm, $\nabla V(\underline{x}^k)$ defines the tangent plane $H(\underline{x} \mid \underline{x}^k) = \{\underline{x} \mid \nabla V(\underline{x}^k)(\underline{x} - \underline{x}^k) = 0\}$ and the halfspace $H^+\{\underline{x} \mid \nabla V(\underline{x}^k)(\underline{x} - \underline{x}^k) > 0\}$. By construction, $\nabla p(\underline{x}^k) = c\nabla V(\underline{x}^k)$ for some positive scalar c, so $H(\underline{x} \mid \underline{x}^k)$ is also the tangent plane to $\nabla p(\underline{x})$ at $\underline{x}^k$. The indifference curves of the proxy are strictly convex,

so if $p(\underline{y}) > p(\underline{x})$ for any $\underline{y} \in X$, then $\underline{y} \in H^{+}(\underline{x} \mid \underline{x}^{k})$. But $\underline{x}^{k+1}$ maximizes $p[\underline{x} \mid \underline{x}^{k}, \underline{x}^{k-1}]$, so $p(\underline{x}^{k+1}) > p(\underline{x}^{k})$. Therefore $\underline{x}^{k+1} \in H^{+}(\underline{x} \mid \underline{x}^{k})$ and $\nabla V(\underline{x}^{k})(\underline{x}^{k+1} - \underline{x}^{k}) > 0$. Q.E.D.

Figure 3.4 illustrates the concept of maximization by proxy. In both the top and bottom diagrams, $\underline{x}^{k+1}$ is the point generated by the k^{th} iteration. The curve labeled *proxy* is the slice of the proxy function along the line determined by $\underline{x}^{k}$ and $\underline{x}^{k+1}$. The quantity $[p(\underline{x}^{k+1}) - p(\underline{x}^{k})]$, measuring the change in the proxy, must be positive (or zero if $\underline{x}^{k+1} = \underline{x}^{k}$) since $\underline{x}^{k+1}$ maximizes the proxy $[p(\underline{x}) \mid \underline{x}^{k}, \underline{x}^{k-1}]$ subject to the constraints. This quantity, represented in Figure 3.4 by the arrow P, is used to predict the improvement in the true objective. Pretending for the moment the true objective is known, we draw the corresponding slice of V along the same direction. The arrow A measures the *actual change* in V at $\underline{x}^{k+1}$. This quantity, $[V(\underline{x}^{k+1}) - V(\underline{x}^{k})]$, can be negative or positive. If the proxy is a good model of the true objective, as drawn in the top diagram, the actual change is positive and the iteration makes an improvement. If the proxy is a poor model, as drawn in the bottom diagram, A and P point in opposite directions. In this case, $\underline{x}^{k+1}$ overshoots the true maximum far enough so $V(\underline{x}^{k+1}) < V(\underline{x}^{k})$. Since movement in the $\underline{x}^{k+1}$ direction guarantees improvement if small enough a step is taken, there must be a point $\underline{x}^{(k+1)'}$ along the line determined by $\underline{x}^{k}$ and $\underline{x}^{k+1}$, but somewhere in between, such that $V(\underline{x}^{(k+1)'}) > V(\underline{x}^{k})$. An optimization procedure called relaxation is used to find such a point $\underline{x}^{(k+1)'}$. For some α, $0 < \alpha < 1$,

$$\underline{x}^{(k+1)'} = \alpha \underline{x}^{k+1} + (1-\alpha)\underline{x}^{k}.$$

The parameter α is decreased until improvement is achieved. Figure 3.5 shows the *actual change* A' , measured at $\underline{x}^{(k+1)'}$, has the same direction as *predicted improvement* P. Since the proxy always specifies a feasible direction of improvement (except at the solution where $\underline{x}^{k+1} = \underline{x}^{k}$), the relaxation procedure is guaranteed to generate the required $\underline{x}^{(k+1)'}$.

At this point we must carefully examine the theoretical differences between the

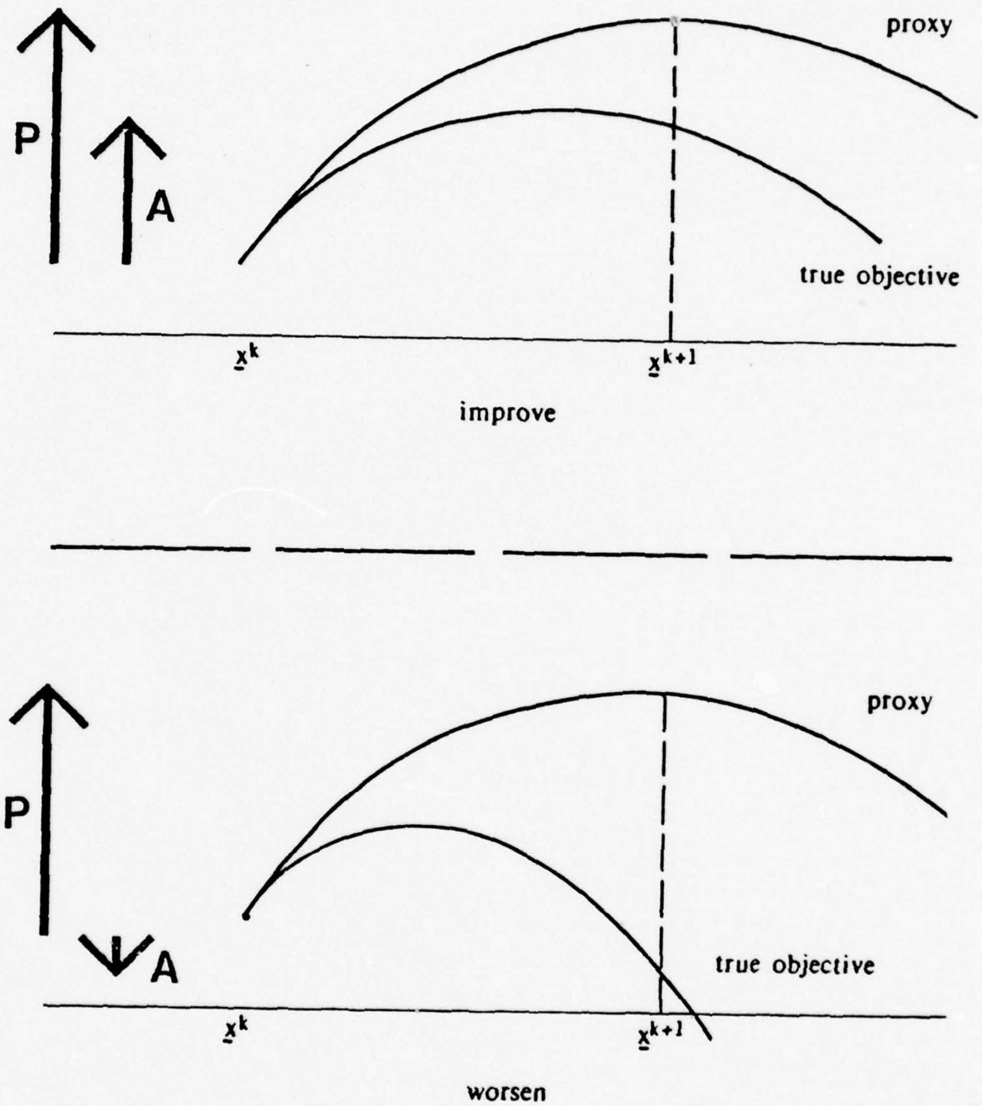


Figure 3.4 Does Iteration Improve or Worsen True Objective?

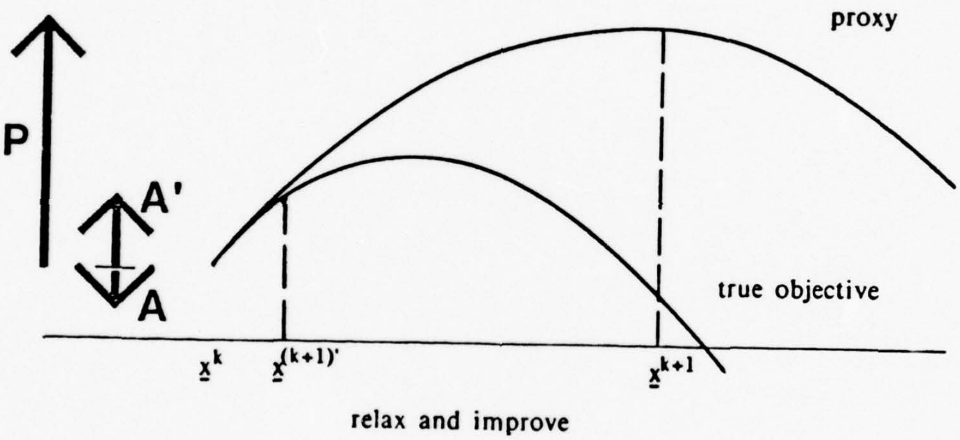


Figure 3.5 Relaxation Technique

modified Frank-Wolfe algorithm and its analog with the sum-of-exponentials proxy. We begin our analysis assuming V is known; then we extend our results to the interactive case where V is unknown.

The original Frank-Wolfe algorithm for maximizing a concave function V over a convex set X is given below:

Step 0. Choose an initial point $\underline{x}^1 \in X$. Let $k = 1$.

Step 1. Determine an optimal solution $\underline{y}^k$ to the direction-finding problem:

$$\max_{\underline{y}} \nabla V(\underline{x}^k)\underline{y} \quad \text{over all} \quad \underline{y} \in X.$$

Let $\underline{m}^k = \underline{y}^k - \underline{x}^k$.

Step 2. Determine an optimal solution t^k to the step-size problem

$$\max_t V(\underline{x}^k + t\underline{m}^k) \quad \text{for} \quad 0 < t < 1.$$

Let $\underline{x}^{k+1} = \underline{x}^k + t^k\underline{m}^k$. Stop if $\underline{x}^{k+1} = \underline{x}^k$. If not, let $k = k+1$ and return to step 1.

Step 1 works directly on the optimality conditions for the true objective V. It finds the best feasible direction of improvement for an infinitesimal step. In step 2, an exact line search is performed, locating the maximum of V in the $\underline{m}^k$ direction. If $t^k = 0$, no feasible direction of improvement exists, so $\underline{x}^k$ is the optimal solution. Wolfe [34] has proven global convergence of this algorithm.

In the modified Frank-Wolfe algorithm, an *inexact* line search is used at step 2. Even when an exact line search is performed, it yields only a trial solution to the original problem. Since exact line search is generally a costly procedure, the inexact line search would be advantageous if convergence could be guaranteed.

Several papers in the optimization literature investigate convergence of feasible direction algorithms. Goldstein [13] devised a technique ensuring finite improvement at each stage of the modified Frank-Wolfe algorithm. At any iteration, the *actual improvement* may be very small even if the *predicted improvement* is large. The true objective, however, must improve by a sufficient amount to prevent jamming at a non-optimal point (jamming occurs when consecutive iterations stall in a non-optimal

level set). The Goldstein test simultaneously tests for improvement and prevents jamming by requiring the ratio of the *actual change* to the *predicted improvement* to exceed a fixed positive ε at each iteration. For the linear approximation algorithm, the test is written mathematically:

If for some fixed $\varepsilon > 0$,

$$\left([V(\underline{x}^{k+1}) - V(\underline{x}^k)] \;/\; [\nabla V(\underline{x}^k)(\underline{x}^{k+1} - \underline{x}^k)] \right) > \varepsilon$$

then $\underline{x}^{k+1}$ is an improvement. If not, let $\underline{x}^{(k+1)'} = \alpha\underline{x}^{k+1} + (1-\alpha)\underline{x}^k$, $0 < \alpha < 1$, and continue decreasing α until the criterion is satisfied. A numerical value for ε must be set as a uniform lower bound on the test ratio.

Armijo [13] designed a very efficient stepsize scheme to accelerate the relaxation procedure. As the trial sequence improves, the technique specifies smaller and smaller steps:

$$\underline{x}^{(k+1)'} = \alpha^n(\underline{x}^{k+1}) + (1-\alpha^n)\underline{x}^k, \quad 0 < \alpha < 1,$$

where n increases by one after each relaxation. The Goldstein test coupled with Armijo relaxation is a very efficient procedure for insuring improvement at each iteration. Garcia-Palomares [13] proved global convergence of the Goldstein and Armijo procedures in linear approximation feasible direction algorithms. His theorem and a formal statement of the Goldstein and Armijo procedures are included in Appendix D; the main concepts of his proof are outlined below:

A. Verify that the algorithm generates a bounded feasible direction of improvement at each iteration.

B. Use the Armijo relaxation procedure to show an improvement satisfying the Goldstein test can be made at each stage.

C. Use continuity and boundedness to prove the limit of the predicted improvement is zero. This step insures the algorithm will converge to some $\underline{x}^\dagger$.

D. Show that any accumulation point $\underline{x}^\dagger$ obeys the optimality criterion $\nabla V(\underline{x}^\dagger)\underline{m}^\dagger \leq 0$ where $\underline{m}^\dagger$ is any feasible direction at $\underline{x}^\dagger$. Thus the algorithm must

converge, and any point to which it converges must be the true optimum.

Now we examine the convergence of the algorithm with the sum-of-exponentials proxy. At each iteration, maximization of the sum-of-exponentials proxy generates a direction $\underline{l}^k$ different than the $\underline{m}^k$ obtained by the Frank-Wolfe procedure. The direction $\underline{l}^k$ is, in a sense, a *global* direction of improvement in contrast to the Frank-Wolfe *local* direction of improvement. This global direction is a better proxy for the path to the true optimum. To prove infinite convergence, however, we must still show that $\nabla V(\underline{x}^\dagger)\underline{m}^\dagger \leq 0$ at any accumulation point $\underline{x}^\dagger$ [13],[32]. The Frank-Wolfe algorithm works directly on this condition, but the sum-of-exponentials algorithm maximizes $p(\underline{x})$ instead.

We may take either of two approaches to ensure global convergence. The first is an extension of the Goldstein-Armijo procedure. It requires an existence proof of a fixed $\delta > 0$, such that for all k,

$$\left([V(\underline{x}^{SE}) - V(\underline{x}^k)] \; / \; [\nabla V(\underline{x}^k)(\underline{x}^{FW} - \underline{x}^k)]\right) > \delta,$$

where $\underline{x}^{SE}$ and $\underline{x}^{FW}$ are the points generated by the sum-of-exponentials and linear proxies, respectively (accounting for relaxation steps if necessary). This existence proof would require a global restriction on the curvature of V. Our proxy approach, however, was designed to avoid global preference assumptions of this type.

A second approach, motivated by the Spacer Step Theorem (see Appendix C), allows us to ensure convergence with no additional assumptions. In this approach, a modified Frank-Wolfe step, using a linear proxy, is inserted periodically between sum-of-exponential iterations. It is called a spacer step since it separates disjoint portions of the complex sequence. The Spacer Step Theorem, when applied to our decision problem, guarantees that if

a) the spacer step is a step of an algorithm known to converge, and

b) all other steps of the process do not worsen the objective,

then

c) the entire complex sequence converges.

Convergence of the modified Frank-Wolfe algorithm (MFW) guarantees (a) is true, and Theorem 3.2 gurarantees (b) is true. Therefore, (c) implies the proxy iteration algorithm with the MFW spacer step is globally convergent.

By making assumptions about the tradeoff assessment error, we can also prove global convergence of the interactive algorithm. These results are discussed together with consistency tests in section 3.5.

Stopping Rule

When interacting with a decision maker, we can assess tradeoffs at only a small number of points, so the number of iterations is limited. Instead of examining the infinite convergence of the sequence $\{\underline{x}^k\}$, we need to calculate a bound on the error $V(\underline{x}^*) - V(\underline{x}^k)$ from stopping at the k^{th} iteration. The strict concavity of V implies

$$V(\underline{x}^*) - V(\underline{x}^k) < \nabla V(\underline{x}^k)(\underline{x}^* - \underline{x}^k).$$

Since $(\underline{x}^* - \underline{x}^k)$ is one of many feasible directions at $\underline{x}^k$,

$$\nabla V(\underline{x}^k)(\underline{x}^* - \underline{x}^k) \le \max_{\underline{y}} \nabla V(\underline{x}^k)\underline{y}$$

where $\underline{y}$ is any feasible direction at $\underline{x}^k$. By construction, $(\partial V(\underline{x}^k)/\partial x_1)\ \lambda(\underline{x}^k)^T = \nabla V(\underline{x}^k)$, so we have an upper bound on the error term:

$$V(\underline{x}^*) - V(\underline{x}^k) \le \max_{\underline{y} \in X(D)} [\partial V(\underline{x}^k)/\partial x_1]\underline{\lambda}(\underline{x}^k)^T \underline{y}.$$

Axiom 2.2, implying the decision maker's willingness to make tradeoffs, guarantees that $\partial V(\underline{x})/\partial x_1$ is bounded above for all $\underline{x} \in X(D)$. Assuming an upper bound κ on this scaling factor, we can calculate an upper bound on the error from stopping at $\underline{x}^k$:

$$V(\underline{x}^*) - V(\underline{x}^k) \le \max_{\underline{y} \in X(D)} \kappa[\underline{\lambda}(\underline{x}^k)^T]\underline{y} \tag{3.3}$$

The optimal $\underline{y}$ maximizes the linear approximation of V, fit at $\underline{x}^k$, subject to the constraints. It is therefore the same point that would be found by a step of the modified Frank-Wolfe procedure. Since the objective $\kappa[\underline{\lambda}(\underline{x}^k)^T]\underline{y}$ is linear, the percentage additional computation for the error bound at each iteration is small. No additional computation is required at the MFW spacer steps.

As a stopping rule for the algorithm, we set a tolerance level ϵ for the error term $V(\underline{x}^*) - V(\underline{x}^k)$. If at iteration k, the error bound in equation (3.3) is less than ϵ, then $\underline{x}^k$ is an acceptable solution. If the error bound exceeds the tolerance, the iterative process continues. This stopping rule is the final step of the algorithm.

For the special case with one linear constraint, the solution to

$$\max_{\underline{y}} \ \underline{\lambda}^T\underline{y} \qquad \text{subject to} \qquad \underline{c}^T\underline{x} \leq b, \ \ \underline{x} > \underline{0}$$

is easily found since the entire b is allocated to the x_j with the highest simplex multiplier. The optimal solution for this subproblem is

$$y_j = b/c_j, \quad y_i = 0, \quad i \neq j, \quad \text{where} \quad j = \max_i \ \lambda_i/c_i.$$

The upper bound on the error after the k^{th} iteration is $\kappa(b/c_j)\lambda_j(\underline{x}^k)$.

Thus far I have developed piece by piece the procedures for assessing tradeoffs, fitting and maximizing the proxy, insuring convergence, and testing trial solutions. Figure 3.6 is a flow chart linking together these individual steps to form the *proxy iteration algorithm*. After the constraints are modeled and the tolerance is established, the analyst assesses the decision maker's tradeoffs. For the sum-of-exponentials proxy, $2N-1$ tradeoffs are required. On the first iteration, $N-1$ tradeoffs at each of two arbitrarily chosen points $\underline{x}^1$ and $\underline{x}^2$ plus one additional tradeoff at a third point nearby provide the necessary information. The sum-of-exponentials proxy is then fit from these assessments and maximized with the techniques of section 3.3. The decision maker must then compare the new maximum $\underline{x}^3$ with the previous point. If $\underline{x}^2 > \underline{x}^3$, relaxation methods are used until $\underline{x}^{3'} > \underline{x}^2$. If the upper bound on $[V(\underline{x}^*) - V(\underline{x}^k)]$ passes the tolerance test, $\underline{x}^{3'}$ is the solution. If it fails, the next iteration begins, but requires only $N-1$ new tradeoff assessments since N tradeoffs at previous points are used again. The iterations continue, using the MFW spacer step periodically, until the error bound at some $\underline{x}^k$ passes the tolerance test; this $\underline{x}^k$ is an acceptable solution. The next $\underline{x}^{k+1}$ could be found with no additional information since $\underline{\lambda}(\underline{x}^k)$ was required for the tolerance test. If $\underline{x}^{k+1} > \underline{x}^k$, $\underline{x}^{k+1}$ would further reduce the error. The tolerance limit can be adjusted to suit the requirements of any particular problem; the

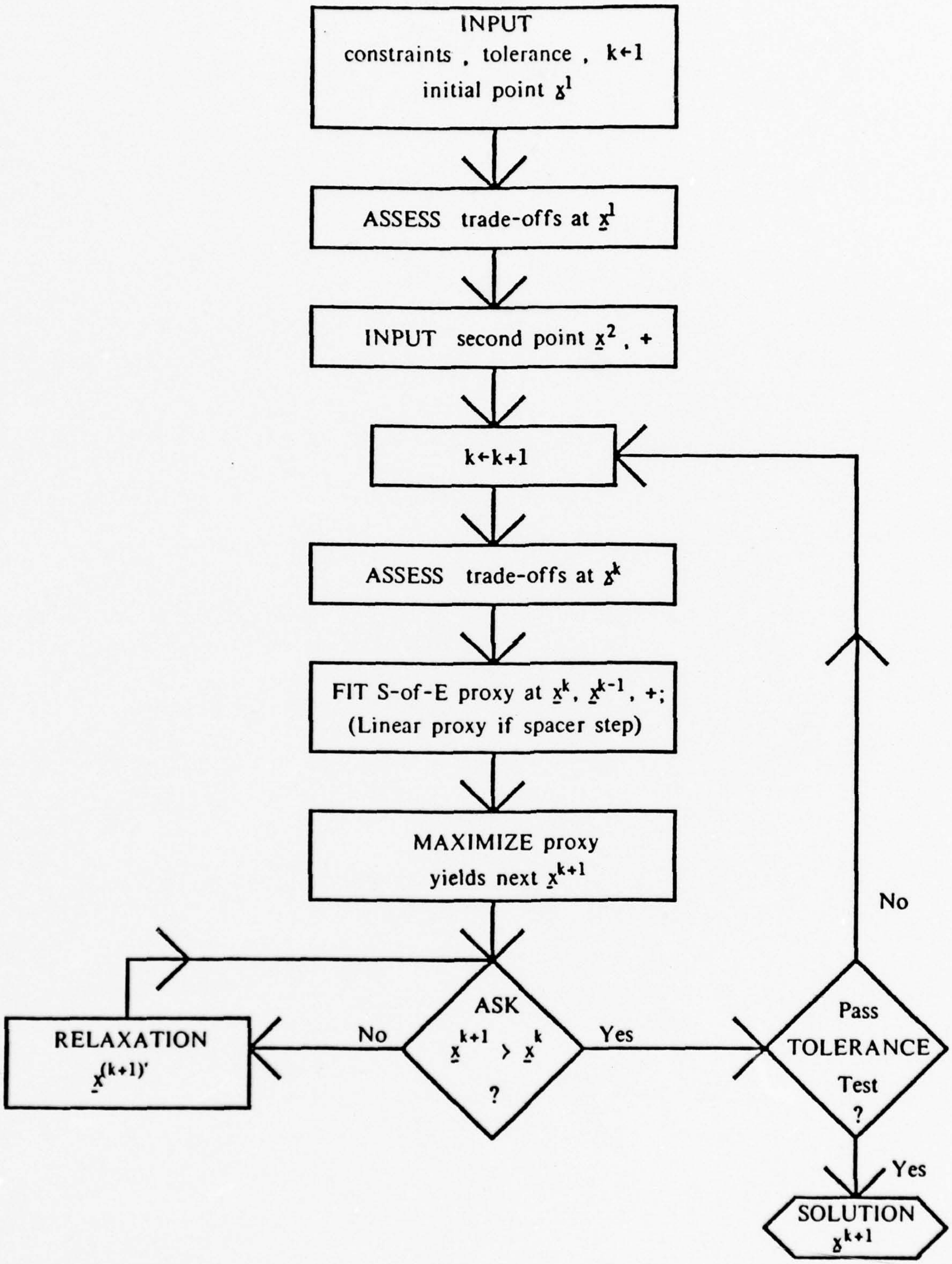


Figure 3.6 Proxy Iteration Algorithm with Sum-of-Exponentials Proxy

smaller the tolerance, the larger the number of iterations required.

Before interacting with a real decision maker, I test the new algorithm with a few simple examples in which the computer plays the role of the decision maker. Given a deterministic preference function, the computer evaluates the tradeoffs, checks for improvement, and performs the tolerance test. There is no assessment error when the computer plays the decision-making role. Figure 3.7 presents this entire computerized procedure.

Although this dissertation is written from a decision analytic perspective, it may be of some interest in optimization as well. Independent of its decision analysis interpretation, the algorithm in Figure 3.7 is itself an optimization technique. Most feasible direction methods use Taylor series approximations, but this algorithm uses sum-of-exponentials, sum-of-powers, and Cobb-Douglass proxies. These functions require less information since they are separable, implying all off-diagonal elements of the inverted Hessian matrix are zero. They are good proxies in our decision problem because of their normative motivation. Perhaps they can be utilized effectively in other search algorithms as well. I have not found any iterative search algorithms in the optimization literature using these functions as local approximations.

Returning now to our decision analysis context, we let the computer play the role of the decision maker. In the following two examples shown in Figures 3.8 and 3.9, the true preference functions are known and closed-form solutions exist. These simple examples are run to illustrate the concepts of this chapter. A different tolerance test may be used since the solutions are known in advance. In Figure 3.7, the iterations stop when each component of $\underline{x}^k$ is within 1% of the corresponding component of $\underline{x}^*$:

$$\text{Stop at } \underline{x}^k \text{ if } \left(|x_j^* - x_j^k| \,/\, x_j^*\right) < 1\% \text{ for all } j.$$

Figure 3.8 shows the interactive nature of the algorithm. The tradeoff vector and new maximum at each iteration are listed until the optimum is reached. The sum-of-exponentials proxy is used in this trivial example. Since the true objective is also a sum-of-exponentials function, the algorithm converges in two steps: $\underline{x}^3$ maximizes the first proxy, and $\underline{x}^4 = \underline{x}^3$ maximizes the second proxy, identical to the

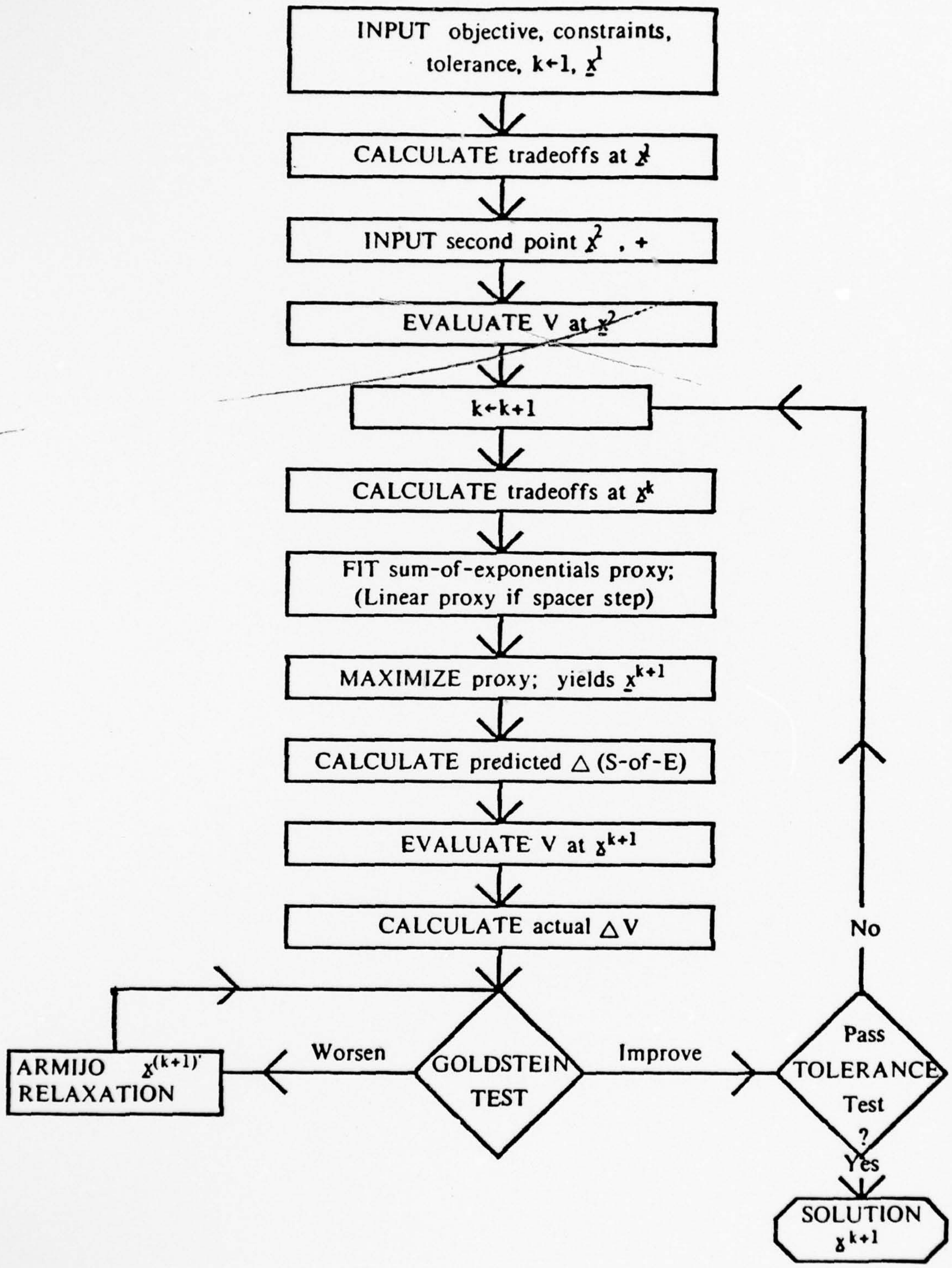


Figure 3.7 Proxy Iteration Algorithm, Given True Objective

$$\text{Maximize} \quad -e^{-0.1x_1} - 4e^{-0.2x_2} - 2e^{-0.4x_3}$$

$$\text{subject to} \quad x_1 + 2x_2 + 4x_3 \leq 100$$

```
ENTER NUMBER OF ATTRIBUTES
⎕:
      3

SELECT DETERMINISTIC PREFERENCE FUNCTION.  TYPE THE NUMBER PRECEDING CHOICE
1 = SUM-OF-EXPONENTIALS;  2 = COBB-DOUGLASS;  3 = POSYNOMIAL;  4 = OTHER
⎕:
      1

ENTER PARAMETERS OF DETERMINISTIC PREFERENCE FUNCTION CHOSEN ABOVE.

ENTER WEIGHTING FACTORS A(1), A(2), ..., A(N)
⎕:
      1 4 2

ENTER EXPONENTS OMEGA(1), OMEGA(2), ... , OMEGA(N)
⎕:
      0.1 0.2 0.4

ENTER LINEAR COST CONSTRAINT COEFFICIENTS C(1),C(2),...,C(N)
⎕:
      1 2 4

ENTER CONSTRAINT MAXIMUM D
⎕:
      100

ENTER TOLERANCE OF SOLUTION AS THE PERCENTAGE ERROR IN EACH COMPONENT
⎕:
      1

ENTER COORDINATES OF INITIAL POINT X(1)
⎕:
      20 10 15

ENTER COORDINATES OF SECOND POINT X(2)
⎕:
      10 40 2.5
```

Figure 3.8a. Example with Sum-of-Exponentials Objective

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(2) AND X(1) YIELDS

NEW MAXIMUM X(3) 26.40 20.13 8.33

PREDICTED IMPROVEMENT IS 0.89094
ACTUAL IMPROVEMENT IS 0.89094

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(3) AND X(2) YIELDS

NEW MAXIMUM X(4) 26.40 20.13 8.33

PREDICTED IMPROVEMENT IS 0.00000
ACTUAL IMPROVEMENT IS 0.00000

OPTIMUM HAS BEEN REACHED. OPTIMAL SOLUTION IS:

26.40 20.13 8.33

LAGRANGE MULTIPLIER AT SOLUTION IS 0.00713

SEQUENCE OF ALGORITHM

POINT			TRADE-OFFS			NUMERAIRE
20.00	10.00	15.00	1.00	8.00	0.15	
10.00	40.00	2.50	1.00	0.01	8.00	⁻1.10498
26.40	20.13	8.33	1.00	2.00	4.00	⁻0.21404
26.40	20.13	8.33				⁻0.21404

Figure 3.8b Iterations with Sum-of-Exponentials Example

$$\text{Maximize} \quad x_1^{0.2}\, x_2^{0.5}\, x_3^{0.3}$$

$$\text{subject to} \quad x_1 + 2x_2 + 4x_3 \leq 100$$

```
ENTER NUMBER OF ATTRIBUTES
□:
      3

SELECT DETERMINISTIC PREFERENCE FUNCTION.  TYPE THE NUMBER PRECEDING CHOICE
1 = SUM-OF-EXPONENTIALS;  2 = COBB-DOUGLASS;  3 = POSYNOMIAL;  4 = OTHER
□:
      2

ENTER PARAMETERS OF DETERMINISTIC PREFERENCE FUNCTION CHOSEN ABOVE.

ENTER EXPONENTS BETA(1), BETA(2), ... , BETA(N)
□:
      0.2 0.5 0.3

ENTER LINEAR COST CONSTRAINT COEFFICIENTS C(1),C(2),...,C(N)
□:
      1 2 4

ENTER CONSTRAINT MAXIMUM D
□:
      100

ENTER TOLERANCE OF SOLUTION AS THE PERCENTAGE ERROR IN EACH COMPONENT
□:
      1

ENTER COORDINATES OF INITIAL POINT X(1)
□:
      20 10 15

ENTER COORDINATES OF SECOND POINT X(2)
□:
      10 40 2.5
```

Figure 3.9a Example with Cobb-Douglass Objective

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(2) AND X(1) YIELDS

NEW MAXIMUM X(3) 16.29 24.74 8.56

PREDICTED IMPROVEMENT IS 0.54992
ACTUAL IMPROVEMENT IS 3.35486

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(3) AND X(2) YIELDS

NEW MAXIMUM X(4) 19.69 24.65 7.75

PREDICTED IMPROVEMENT IS 0.01058
ACTUAL IMPROVEMENT IS 0.10683

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(4) AND X(3) YIELDS

NEW MAXIMUM X(5) 20.19 24.71 7.59

PREDICTED IMPROVEMENT IS 0.00028
ACTUAL IMPROVEMENT IS 0.00288

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(5) AND X(4) YIELDS

NEW MAXIMUM X(6) 19.81 25.19 7.45

PREDICTED IMPROVEMENT IS 0.00020
ACTUAL IMPROVEMENT IS 0.00058

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

SUM OF EXPONENTIALS FITTED LOCALLY FROM X(6) AND X(5) YIELDS

NEW MAXIMUM X(7) 20.00 25.00 7.50

PREDICTED IMPROVEMENT IS 0.00006
ACTUAL IMPROVEMENT IS 0.00051

ITERATION IMPROVED OBJECTIVE. USE NEW MAXIMUM IN NEXT ITERATION.

OPTIMUM HAS BEEN REACHED. OPTIMAL SOLUTION IS:

20.00 25.00 7.50

Figure 3.9b Iterations with Cobb-Douglass Example

LAGRANGE MULTIPLIER AT SOLUTION IS 0.01836

SEQUENCE OF ALGORITHM

POINT			*TRADE-OFFS*			*NUMERAIRE*
20.00	10.00	15.00	1.00	5.00	2.00	
10.00	40.00	2.50	1.00	0.63	6.00	13.19508
16.29	24.74	8.56	1.00	1.65	2.86	16.54994
19.69	24.65	7.75	1.00	2.00	3.81	16.65677
20.19	24.71	7.59	1.00	2.04	3.99	16.65965
19.81	25.19	7.45	1.00	1.97	3.99	16.66023
20.00	25.00	7.50				16.66074

Figure 3.9c Solution of Cobb-Douglass Example

first. The predicted improvement is equal to the actual improvement in both iterations; both are zero at the solution. Figure 3.8b shows the optimal solution and its associated shadow price in terms of the numeraire $V(\underline{x})$.

The example in Figure 3.9 is more interesting since the sum-of-exponentials proxy is used to optimize a Cobb-Douglass true objective. The tolerance test is passed after five iterations. Each new maximum is an improvement, so no relaxations are required. The computer solution agrees with the analytic solution to seven decimal places. In these examples, the period of the spacer step was ten; only five iterations were required, so the MFW step was never used. In the next chapter, this same problem is solved with Boyd's algorithm and the rates of convergence are compared.

Nonconvex Feasible Regions

In proving Theorem 3.1, we assumed that the set of feasible alternatives was convex. If $X(D)$ is not convex, the maximization of $p[\underline{x}(\underline{d}) \mid \underline{x}(\underline{d}^*), \underline{x}(\underline{d}^k)]$ at $\underline{x}(\underline{d}^*)$ is not a necessary condition for the maximization of $V(\underline{x})$ since there may be no hyperplane that separates the constraint and indifference surfaces at $\underline{x}(\underline{d}^*)$. The condition is still sufficient if $X(D)$ is a connected set, but there is no guarantee it will obtain. The convexity assumption prohibits gaps and discrete outcomes. In the language of production functions, it corresponds to constant or decreasing returns to scale. Theorem 3.1 is consistent with the Arrow-Hurwicz result [2] that pricing systems allocate resources efficiently when there are no increasing returns to scale. Convexity is a plausible assumption, applicable to many multi-attribute decisions with continuous outcome variables.

The proxy iteration algorithm may still be useful in cases where $X(D)$ is not convex. If $X(D)$ is locally convex in the region of the optimum, maximization of the proxy at $\underline{x}(\underline{d}^*)$ is still a necessary and sufficient condition for the maximization of $V(\underline{x})$. If the constraint has a gap near the optimum, the algorithm may be applied to its convex envelope. If the gap is small, sensitivity analysis may reveal the true optimum after the solution to this artificial problem is found. Discrete outcome decisions present

a greater obstacle since small steps in a direction of improvement may be infeasible when a large step is allowed. Consequently, the algorithm could stop at a non-optimal point. Traditional multi-attribute procedures should be used for discrete problems that have no meaningful continuous analogs.

3.5 Consistency of Tradeoff Assessments

We have assumed up to this point that at each iteration the decision maker provides tradeoffs consistent with a continuously differentiable deterministic preference function. This assumption has allowed us to develop the proxy iteration algorithm without complications arising from assessment error. In this section, we relax the assumption and examine techniques for checking tradeoff consistency. We also examine the effects of assessment error on the convergence of the algorithm.

I view the proxy iteration algorithm as a learning process for the decision maker. At each iteration, the decision maker learns more about his underlying preferences as he sees the implications of his previous assessments. He benefits not only from this feedback, but also gets more practice interpreting and responding to the analyst's questions at each step. Viewing the interactive algorithm as a learning process, we can realistically assume that the tradeoff assessments become better representations of the decision maker's preferences as the iterations proceed. Any cumulative consistency checking scheme should embody this idea.

I augment the algorithm in Figure 3.6 with two types of consistency tests, the first testing tradeoff consistency at a single point, and the second testing consistency at successive points. The single point test is a standard procedure described in almost every tradeoff assessment scheme in the literature [4],[14]. It requires a second set of assessments at each point, using a different price variable x_k. For price variable x_1, $\lambda_{1j} = dx_1/dx_j$; for price variable x_k, $\lambda_{kj} = dx_k/dx_j$. The chain rule implies $\lambda_{1k} \equiv \lambda_{1j}\,\lambda_{jk}$. Since only N-1 unique tradeoffs among the attributes exist at any point, the second set can be used to measure the discrepancy:

$$\%\ \text{error} = \left[(\Delta x_1/\Delta x_k) - (\Delta x_1/\Delta x_j)\,(\Delta x_j/\Delta x_k)\right] \Big/ \left(\Delta x_1/\Delta x_k\right)$$

Certainly we would not expect exact agreement. Instead, we set a reasonable tolerance level; if the discrepancy exceeds the tolerance, the analyst should explain the inconsistency to the decision maker and reassess the tradeoffs until the discrepancy is resolved.

This scheme checks tradeoff consistency at a single outcome, but does not reveal information about the shape of the indifference curves. Axiom 2.4 requires strictly convex indifference curves since the marginal rates of substitution are decreasing. To check for violations of this axiom, tradeoffs at outcomes on the same indifference curve must be compared. Our algorithm, however, never generates outcomes on the same indifference curve since each iteration yields an improvement. Since our primary motivation is the reduction of the total number of assessments required to reach the optimum, we certainly do not want to make many extra assessments just to check consistency. We must try to strike a balance between the number of assessments required to reach a given consistency level and the level of consistency required to reach the optimum in a small number of iterations. If the tradeoff assessments are very inconsistent, the proxies will be poor preference models and the resulting trial sequence will converge very slowly at best.

Fortunately, the algorithm in Figure 3.6 provides a consistency check for tradeoffs at successive points without requiring any extra assessments. It is easy to prove that the sum-of-exponentials proxy obeys Axiom 2.4 if and only if the parameters $\underline{a}$ and $\underline{\omega}$ are strictly positive (see Appendix B; analogous results hold for the Cobb-Douglass and sum-of-powers proxies as well). When the proxy is fit from the current and previous tradeoffs at each iteration, the sign of each a_i and ω_i can be checked. We must realize that this scheme checks for decreasing marginal rates of substitution of the proxy only. The indifference curves of the true objective are unknown and their convexity cannot be verified directly without numerous additional assessments. If any a_i or ω_i is negative or zero, the tradeoffs should be reassessed. If the parameter remains nonpositive, the algorithm could continue with a MFW spacer step using just the tradeoffs at the current point. Figure 3.10 shows the proxy iteration algorithm with both consistency tests.

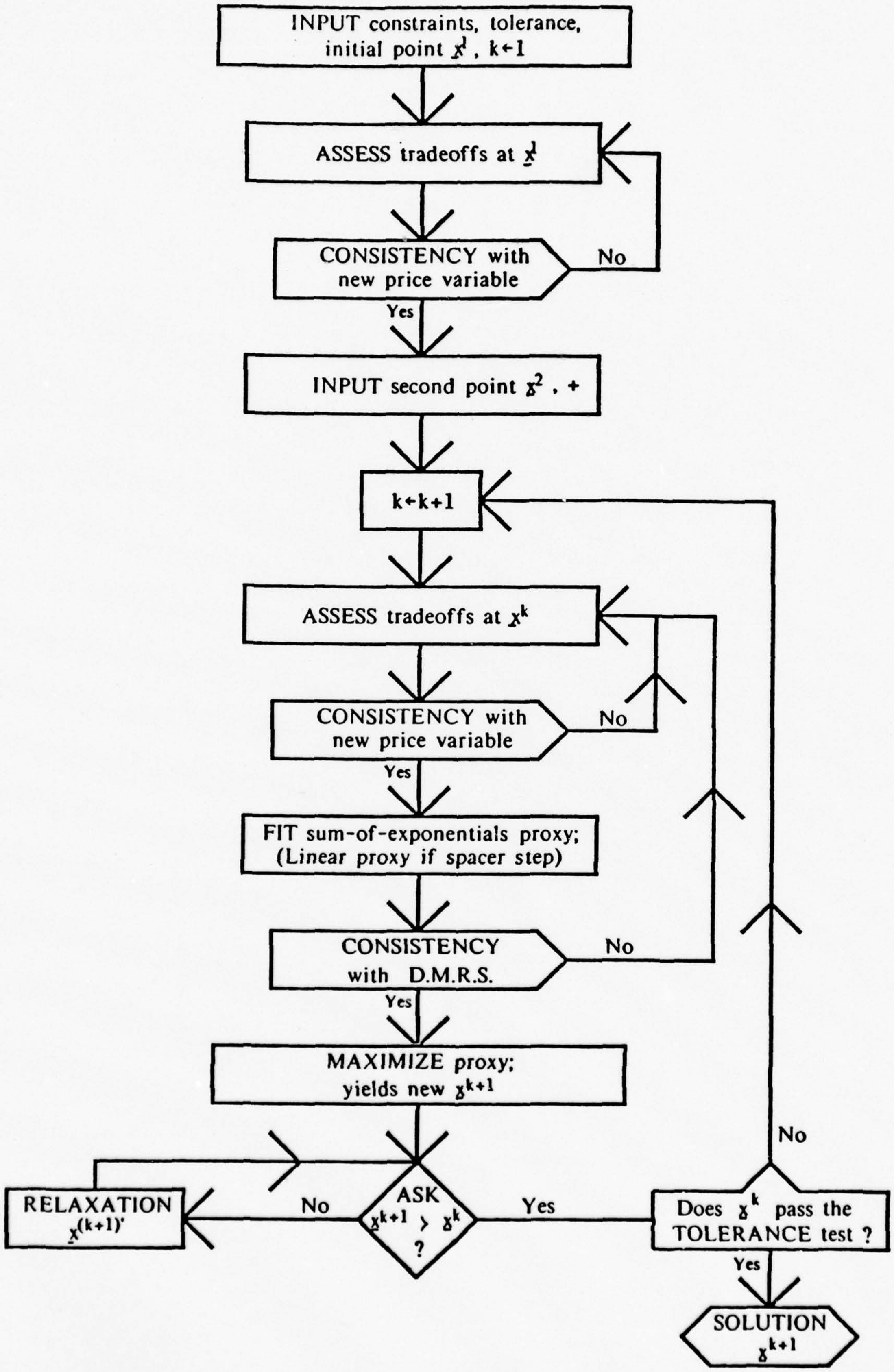


Figure 3.10 Proxy Iteration Algorithm with Consistency Tests

Although I do not suggest a least squares approach, it could be used to handle the consistency problem. Instead of fitting the proxy exactly from $2N-1$ assessments, the parameters could be chosen to minimize the error from all or any subset of previous assessments, assigning the heaviest relative weights to the most recent tradeoffs. Initially, the idea of utilizing all previous information sounds promising, but after setting up the problem, we see that this nonlinear least squares fit is a very difficult optimization problem. For the sum-of-exponentials proxy, we must minimize E^2 with respect to $\omega_1, \omega_2, \ldots, \omega_N$ and $a_2, a_3, \ldots, a_N$ at each iteration, where

$$E^2 = \sum_{k=1}^{K} \sum_{i=2}^{N} \left(\lambda_i(\underline{x}^k) - [(\omega_i a_i)/\omega_1]\, e^{\omega_1 x_1{}^k - \omega_i x_i{}^k}\right)^2.$$

The first index sums over all previous assessments. Quasi-Newton methods or successive linear approximation on the normal equations, even with least squares updating procedures, would require a huge additional amount of computer time at each iteration. Even if we ignored the computation cost, we may still have poor parameter fits since each a_i counterbalances the effect of its associated ω_i.

This discussion of the least squares approach reminds us that our primary goal is finding the optimum while incurring a reasonably small assessment and computation cost. Complex time-consuming consistency checks should not be used since we still have only a proxy at each iteration regardless of the effort expended. The two consistency checks I employ, each requiring minimal extra assessment and computation, should be sufficient to keep the trial sequence from going astray.

Research in a similar vein at U.C.L.A. supports this conclusion. Geoffrion, Dyer, and Feinberg [14],[8] and Hogan [15] developed an interactive Frank-Wolfe algorithm. Hogan investigated the effects of error in the gradient when using the modified Frank-Wolfe method. At each iteration, he added a noise vector $\underline{\eta}$ to the true gradient $\nabla V(\underline{x})$. At the k^{th} iteration, his noisy gradient $\underline{z}^k$ is the sum of the true gradient and the error term,

$$\underline{z}^k = \nabla V(\underline{x}^k)^T + \underline{\eta}^k.$$

Hogan [15] proved that if $\lim_{k\to\infty} \underline{\eta}^k = \underline{0}$, then for any infinite sequence $\{\underline{x}^k\}$

generated by the algorithm, every accumulation point $\underline{x}^{\dagger}$ is a solution. The assumption $\underline{\eta}^k \to \underline{0}$ implies in our decision-making context that the assessment procedure is a learning process with diminishing error. Dyer [9] extended Hogan's result for the case where $\underline{\eta}^k \to \underline{\eta}^{\infty} \neq \underline{0}$. At each iteration, he replaced $\nabla V(\underline{x}^k)$ by $\alpha^k \nabla V(\underline{x}^k) + \underline{\eta}^{kT}$, where α^k is the scaling factor and $\underline{\eta}^k$ is the assessment error. Dyer defined the approximate solution set

$$\Gamma(\delta) = \{\underline{x} \in X(D) \mid V(\underline{x}) > V(\underline{y}) - \delta \quad \forall \underline{y} \in X(D)\}$$

and proved that if

a) $\lim_{k\to\infty} \alpha^k = \alpha^{\infty}$ and $\lim_{k\to\infty} \underline{\eta}^k = \underline{\eta}^{\infty}$, where $0 < \|\alpha^{\infty}\| < \infty$ and $\|\underline{\eta}^{\infty}\| < \infty$,

then either

b) the modified Frank-Wolfe algorithm terminates at some finite iteration k and $\underline{x}^k \in \Gamma[(\xi^k/\alpha^k)\|\underline{\eta}^k\|]$, where $\xi^k = \max_{\underline{x}\in X(D)} \|\underline{x} - \underline{y}^k\|$, and $\underline{y}^k = \max_{\underline{y}\in X(D)} \alpha^k \nabla V(\underline{x}^k)\underline{y} + \underline{\eta}^{kT}\underline{y}$,

or

c) the MFW algorithm generates an infinite sequence $\{\underline{x}^k\}$, and every accumulation point $\underline{x}^{\infty}$ of $\{\underline{x}^k\}$ is contained in $\Gamma[(\xi^{\infty}/\alpha^{\infty})\|\underline{\eta}^{\infty}\|]$, where $\xi^{\infty} = \max_{\underline{x}\in X(D)} \|\underline{x} - \underline{y}^{\infty}\|$, and $\underline{y}^{\infty} = \lim_{k\to\infty} \underline{y}^k$.

Appendix E contains the full statement of Dyer's result. The quantity $[(\xi^k/\alpha^k)\|\underline{\eta}^k\|]$ bounds the error at any iteration k; it is the analog of equation (3.3), adjusted for the assessment error $\underline{\eta}^k$. As long as the assessment error is bounded, the modified Frank-Wolfe algorithm converges to a solution whose error is bounded. In this case, actual computation of the error bound at iteration k requires specification of an upper bound on $\|\underline{\eta}^k\|$. The Spacer Step Theorem guarantees that Dyer's error analysis holds for the sum-of-exponentials algorithm with the MFW spacer step.

Geoffrion, Dyer, and Feinberg [14] applied their interactive Frank-Wolfe procedure to the allocation of faculty time among teaching, research, and departmental

duties at the U.C.L.A. Graduate School of Management. They claim the decision makers were able to provide the required information "without significant difficulty". In a follow-up paper [9], Dyer concludes:

> Thus the use of interactive programming is not dependent on the assumption of subject behavior consistent with the existence of a preference relation weakly ordering all alternatives. Rather, responses that reflect the 'human element' in the form of random errors and inconsistencies does not appear to be a significant hindrance to the use of this robust procedure.

This chapter concludes the theoretical development of the proxy iteration algorithm under certainty. In the next chapter, I compare the normatively motivated proxies to the linear proxy in a variety of examples, letting the computer play the role of the decision maker. I also state results on the initial rates of convergence of the algorithms.

CHAPTER IV

THE NEW ALGORITHM VERSUS THE OLD

4.1 An Example With Boyd's Algorithm

To see if the new algorithm is faster, I also programmed Boyd's technique and used it to solve the example in Figure 3.9. The trial sequence, beginning at the same $\underline{x}^1$, is listed in Table 4.1. The algorithm generated twenty-six trial points before converging. The list is long because the linear pseudo-objective is a poor preference model; the first trial point at each iteration is an extreme point clearly inferior to $\underline{x}^1$. After observing Table 4.1, Boyd commented that if the first step of each iteration had been used only to guide the search, and not as a trial point, assessments would have been required at only seventeen points. In contrast, the algorithm with the sum-of-exponentials proxy found a more precise solution to this same problem requiring assessments at only six points.

Assessment cost and computation cost are the two criteria to be considered when comparing the algorithms. The total assessment requirement is the key factor since it is by far the most time-consuming and most costly part of the procedure. In the example of Figure 3.9, Boyd's algorithm required assessments at three times as many points as the new algorithm. With respect to the second criterion, Boyd's method used three times as many CPU seconds as well. Although the new technique seems far superior in this example, no general conclusions can be drawn. In the next section, I systematically compare the efficiency of the different proxies.

4.2 Comparison of the New and Old Proxies

With the computer again playing the role of the decision maker, three proxies are tested for three objectives with three different constraint sets. Since the true objectives are given, the optimal solution for each problem can be determined beforehand, and the stopping criterion of Figures 3.8 and 3.9 can be used:

$$\text{Stop at } \underline{x}^k \text{ if } \left(|x_j^* - x_j^k| \;/\; x_j^*\right) \leq 1\% \text{ for all } j.$$

Table 4.1

Trial Sequence of Boyd's Algorithm

with Cobb-Douglass Example

POINT			NUMERAIRE
20.00	10.00	15.00	12.97279
0.05	49.95	0.01	1.04264
16.01	17.99	12.00	15.56594
0.05	49.95	0.01	1.04264
12.82	24.38	9.60	16.21217
99.90	0.02	0.01	0.10665
30.23	19.51	7.69	16.10451
16.30	23.41	9.22	16.46474
99.90	0.02	0.01	0.10665
33.02	18.73	7.38	15.86509
19.65	22.47	8.85	16.54230
0.05	49.95	0.01	1.04264
15.73	27.97	7.08	16.51001
18.86	23.57	8.50	16.60035
0.05	49.95	0.01	1.04264
15.10	28.85	6.80	16.42956
18.11	24.63	8.16	16.62568
99.90	0.02	0.01	0.10665
34.47	19.71	6.53	15.82191
21.38	23.64	7.83	16.63563
0.05	49.95	0.01	1.04264
17.11	28.90	6.27	16.45580
20.53	24.69	7.52	16.65896
0.05	49.95	0.01	1.04264
16.43	29.75	6.02	16.35724
19.71	25.71	7.22	16.65357
20.36	24.90	7.46	16.66006

For each proxy, we want to compare the number of trial points generated before the tolerance test is passed. To give Boyd's algorithm its best performance, the first point at each iteration is used only to specify the direction; it is not counted as a trial point.

Table 4.2 shows the results for the linear, Cobb-Douglass, and sum-of-exponentials proxies in problems with one linear constraint. In order to cover a wide range of problems, representative of those likely to be encountered in practice, the experiments are performed for each proxy, for both symmetric and skew sum-of-exponentials, Cobb-Douglass, and sum-of-powers objectives. In a symmetric objective, all attributes have approximately equal importance; in a skew objective, some attributes are weighted more heavily than others. To avoid any special properties of low dimension, the programs are run for both three- and six-attribute problems.

In all thirty-six examples of Table 4.2, the maximizations at each iteration have closed-form solutions. Maximizing the linear proxy with one linear constraint is a special linear programming problem. The optimum can be found analytically since it is the extreme point with the highest simplex multiplier. For the strictly concave proxies, the primal-dual technique of section 3.3 provides an analytic solution. The computer runs are fast since the closed-form maximizations are programmed directly into the algorithm.

Table 4.2 shows the number of trial points and the number of CPU seconds required for convergence at the 1% tolerance level. It would be unfair to compare Boyd's linear proxy to a concave proxy that was identical to the true objective. The results using the sum-of-exponentials proxy for the sum-of-exponentials objective and the Cobb-Douglass proxy for the Cobb-Douglass objective serve only as reference points. In each example of Table 4.2, the concave proxies are much more efficient than the linear proxy. For the sum-of-exponentials objective, Boyd's linear proxy uses 50% more assessments than the Cobb-Douglass proxy in three dimensions, and 500% to 800% more in six dimensions. For the three-dimensional Cobb-Douglass objective, Boyd's proxy requires five times as many assessments as the sum-of-exponentials proxy. In the six-dimensional problem, I stopped the linear algorithm at 4% and 5% tolerance levels

Table 4.2 Comparison of Proxy Functions

Constraint: Linear (1)

Tolerance: 1%

PREFERENCE FUNCTION		PROXY: LINEAR		PROXY: COBB-DOUGLASS		PROXY: SUM-of-EXPO.	
dimension	shape	# pts. req. assessment	CPU seconds	# pts. req. assessment	CPU seconds	# pts. req. assessment	CPU seconds
Sum-of-Expo.							
3	symmetric	18	2.8	12	1.6	3	0.6
3	skew	14	2.4	10	1.3	3	0.7
6	symmetric	76	12.6	12	1.7	3	0.7
6	skew	71	10.9	8	1.1	3	0.7
Cobb-Douglass							
3	symmetric	22	3.1	2	0.3	5	1.3
3	skew	28	4.4	2	0.3	5	1.1
6	symmetric	45+ 4% Tol.	7.2	2	0.3	7	1.6
6	skew	61+ 5% Tol.	9.8	2	0.3	8	2.2
Sum-of-Powers							
3	symmetric	5 excellent starting pt.	0.7	5 excellent starting pt.	0.8	4 excellent starting pt.	1.0
3	skew	24	3.6	7	1.1	8	2.4 Nonneg.
6	symmetric	44+ 5% Tol.	7.5	10	1.9	6	1.5
6	skew	72+ 2% Tol.	11.4	6	1.0	6	1.6

since it had already used seven times as many assessments and was improving very slowly. The sum-of-powers example allows us to compare all three proxies at once. Even though the Cobb-Douglass function is a special case of the sum-of-powers, the sum-of-exponentials proxy is just as good or better since it is fit from twice as much information. The Cobb-Douglass proxy requires N-1 parameters from one point, whereas the sum-of-exponentials proxy requires 2N-1 tradeoffs from three points. The convergence results with the sum-of-powers objective follow the same pattern, demonstrating the superiority of the normatively motivated proxies. Their small assessment requirement and computation cost make the new algorithm practical, whereas the assessment demand with the linear proxy is prohibitive.

A few basic principles can be drawn from the these results. First and foremost, we should select as our proxy the function that best represents the decision maker's true preferences. Secondly, we should select as our starting point our best estimate of the true optimum. Both of these tasks can be accomplished by using Barrager and Keelin's global assessment procedures, reviewed in section 2.2, as the first step of the algorithm. *In this first step, we encode a global deterministic preference function. Then we begin the iterative procedure, using the global function as the local proxy and its optimum as the starting point. With this combined method, we fully exploit the power of the global procedure, yet avoid its severe restrictions.* We never assume, even in the small, that the proxy is the true objective; we merely use it as a mechanism to guide the search to the optimal solution.

Table 4.3 shows a set of experiments comparing the linear and Cobb-Douglass proxies in problems with the nonlinear constraint $\Sigma_i \, c_i x_i^2 \leq b$. In each problem, the normatively motivated proxy outperforms the linear proxy by a wide margin. Though the following observation does not concern us directly, it is interesting to note that the linear approximation converged faster with the ellipsoidal constraint than with the linear constraint since the initial trial solutions at each iteration were not outlying extreme points.

The third and final set of experiments compares Boyd's pseudo-objective with the sum-of-exponentials proxy in problems with several linear constraints. Linear

Table 4.3 Second Comparison of Proxy Functions

Constraint: Ellipsoidal (1)

Tolerance: 1%

PREFERENCE FUNCTION		PROXY: LINEAR		PROXY: COBB-DOUGLASS	
dimension	shape	# pts.req. assessment	CPU seconds	# pts.req. assessment	CPU seconds
Sum-of-Expo.					
3	symmetric	26	3.6	8	1.2
3	skew	28	3.8	13	1.9
6	symmetric	29	3.9	9	1.2
6	skew	27	3.6	17	2.6
Cobb-Douglass					
3	symmetric	18	2.3	2	0.3
3	skew	23	3.1	2	0.3
6	symmetric	21	2.8	2	0.3
6	skew	27	3.6	2	0.3
Sum-of-Powers					
3	symmetric	11	1.6	3	0.4
3	skew	17	2.5	3	0.4
6	symmetric	13	1.9	4	0.7
6	skew	15	2.2	4	0.7

programming and convex programming routines were required since no analytic solutions exist for these problems. The results in Table 4.4 again show the new proxy is much more efficient than the old, requiring only a small fraction of the assessments and CPU time.

In an attempt to find a case in which Boyd's algorithm was faster, I contrived the following example:

$$\text{Maximize}_{\underline{x}} \quad V(\underline{x}) \;=\; 8x_1^{0.8} + x_2^{0.1} + x_3^{0.1}$$

$$\text{subject to} \quad \Sigma_i\, x_i \;\leq\; 100, \quad \underline{x} \geq \underline{0}.$$

The results with the three proxies are listed below:

Linear:	2 iterations	0.3 CPU seconds
Cobb-Douglass:	2 iterations	0.4 CPU seconds
Sum-of-exponentials:	4 iterations	0.9 CPU seconds.

The optimal solution, $\underline{x}^* = (100, 0, 0)$, is obvious by inspection since the objective is nearly lexicographic. Linear and lexicographic preference structures would be identified early in an analysis and would be modeled with other techniques.

If we formally analyze the rate of convergence, our primary concern is the initial rather than the asymptotic rate since only a modest number of interactive iterations can be performed. Very little is known about the initial rate of convergence of mathematical programming algorithms in general, but for the Frank-Wolfe algorithm, Wolfe [34] and Amor [1] demonstrated the following result:

> If V is boundedly concave (i.e., V is concave with continuous second derivatives on X and a uniform lower bound on all eigenvalues of the Hessian) and X is a convex, compact set, then the error in the objective function is at least halved for each of the first K iterations:
>
> $$\left([V(\underline{x}^*) - V(\underline{x}^{k+1})] \;/\; [V(\underline{x}^*) - V(\underline{x}^k)]\right) \;\leq\; \tfrac{1}{2}, \quad k \leq K,$$
>
> but K remains unknown.

Wolfe claims this result should hold as long as $\underline{x}^K$ is "sufficiently far" from the

Table 4.4 Third Comparison of Proxy Functions

Constraints: Linear (3)

Tolerance: 1%

PREFERENCE FUNCTION		PROXY: LINEAR		PROXY: SUM-of-EXPO.	
dimension	shape	# pts. req. assessment	CPU seconds	# pts. req. assessment	CPU seconds
Cobb-Douglass					
4	skew	12+ 5% Tol.	6.70	4	1.95
Sum-of-Powers					
4	symmetric	21+ 7% Tol.	11.51	3	1.67

optimum. However, the result is not useful since "sufficiently far" is undefined and K is unknown.

Dyer [9] gives the assessment errors $\underline{\eta}^k$ a stochastic interpretation. He proves that if each $\underline{\eta}^k$ is a sample from a multivariate density function with mean $\underline{0}$ and finite variance, then the mean initial rate of convergence of the modified Frank-Wolfe algorithm is unchanged. The Spacer Step Theorem insures the same initial rate of convergence holds for each of the first K groups of steps of the proxy iteration algorithm, but K again remains unknown.

In our proxy approach, Axioms 2.1-2.4 are the only restrictions we wish to impose on the objective function. Even if we make further assumptions bounding the local curvature of V, we still have an infinite set of possibilities. Establishing more concrete results is therefore a very difficult task. Amor [1] summarizes his mathematical analysis claiming *"the initial rate of convergence is primarily determined by the initial* $\underline{x}^0$ *and the relationship between the directional derivatives and the local curvature of* V." His claim is entirely consistent with our intuition and the results we drew from the experiments of this chapter. Since the normatively motivated proxies approximate the local curvature of $V(\underline{x})$ much better than does the linear proxy, they generate a greater improvement at each iteration and yield a higher rate of convergence.

CHAPTER V

THE PROXY ITERATION ALGORITHM FOR DECISION MAKING UNDER UNCERTAINTY

5.1 The Proxy Approach Under Uncertainty

Figure 1.1 shows that under uncertainty, the decision problem is

$$\max_{\underline{d} \in D} \int_{\underline{s}} u[\underline{x}(\underline{d},\underline{s})] \ \{\underline{s}|\varepsilon\}$$

where u is a cardinal utility function and $\{\underline{s}|\varepsilon\}$ is the joint probability distribution over the state variables. Using the preference decomposition approach illustrated in Figure 2.1, the decision problem becomes

$$\max_{\underline{d} \in D} \int_{\underline{s}} u\{n[V(\underline{x}(\underline{d},\underline{s}))]\} \ \{\underline{s}|\varepsilon\}$$

where n is an appropriately chosen numeraire. We are no longer trying to find the most preferred outcome $\underline{x}^*$; rather, we are searching for the most preferred decision $\underline{d}^*$. Therefore, the problem can no longer be structured as a choice of $\underline{x}$ over $X(D)$, but rather as a choice of $\underline{d}$ over D.

Axioms 2.1-2.9 guarantee the existence of a real-valued risk preference function u and deterministic preference function V. Just as under certainty, we assume that the global assessment procedures for u and V are too restrictive, so we try to use our proxy approach instead.

When applying the proxy technique to this problem, we decompose deterministic and risk preferences, first assessing deterministic tradeoffs, then selecting a numeraire and assessing risk preference. We are trying to find the decision $\underline{d}^*$ that maximizes the expected utility $\langle u\{n[V(\underline{x}(\underline{d},\underline{s}))]\}|\varepsilon\rangle$. This composite function notation is cumbersome, so we define the following abbreviated notation showing only the dependence of $u(n)$ upon the decision variables and state variables:

$$u(\underline{d},\underline{s}) \quad = \quad u\{n[V(\underline{x}(\underline{d},\underline{s}))]|\varepsilon\}$$

Taking the expectation with respect to $\underline{s}$,

$$\langle u(\underline{d})\rangle = \int_{\underline{s}} u\{n[V(\underline{x}(\underline{d},\underline{s}))]\}\ \{\underline{s}|\varepsilon\}$$

We also define $p[\underline{d}|\underline{d}^k,\underline{s}]$ and $q[\underline{d}|\underline{d}^k,\underline{s}]$ as approximations of $V(\underline{d},\underline{s})$ and $u(\underline{d},\underline{s})$, respectively, fit from deterministic and risk preference assessments at $\underline{x}(\underline{d}^k,\underline{s})$. Similarly, the notation $\langle q(\underline{d})|\underline{d}^k\rangle$ represents $\int_{\underline{s}} q\{n[p[\underline{x}(\underline{d},\underline{s})|\underline{d}^k]]\}\ \{\underline{s}|\varepsilon\}$. The following theorem is the analog of Theorem 3.1 for decision making under uncertainty.

Theorem 5.1. *If the decision maker's preference ordering satisfies the deterministic and risk preference axioms (2.1-2.9), if* D *is convex and* $\langle u(\underline{d})\rangle$ *is concave, if* $\underline{d}^*$ *is a regular point of the constraints, and if* $q[\underline{d}|\underline{d}^*,\underline{s}]$ *is constructed so* $\nabla_{\underline{d}}\langle q(\underline{d}^*)|\underline{d}^*\rangle = t\nabla_{\underline{d}}\langle u(\underline{d}^*)\rangle$ *for some positive scalar* t, *then if* $\underline{d}^*$ *maximizes* $\langle q[\underline{d}|\underline{d}^*,\underline{s}]\rangle$ *over all* $\underline{d} \in D$, *then* $\underline{d}^*$ *also maximizes* $\langle u(\underline{d})\rangle$ *over all* $\underline{d} \in D$.

The proof of Theorem 5.1 parallels the proof of Theorem 3.1. We write the constraint set D as $D = \{\underline{d} \mid \underline{h}(\underline{d}) = \underline{0},\ \underline{g}(\underline{d}) \leq \underline{0}\}$.

Proof: The regularity conditions hold for all $\underline{d} \in D$, so if $\underline{d}^*$ maximizes $\langle q[\underline{d}|\underline{d}^*,\underline{s}]\rangle$ over all $\underline{d} \in D$, the Kuhn-Tucker necessary conditions guarantee the existence of $\underline{\lambda}$ and $\underline{\mu}$, $\underline{\mu} \geq 0$, such that

$$\nabla_{\underline{d}}\langle q(\underline{d}^*)|\underline{d}^*\rangle + \underline{\lambda}\nabla_{\underline{d}}\underline{h}(\underline{d}^*) + \underline{\mu}\nabla_{\underline{d}}\underline{g}(\underline{d}^*) = \underline{0}.$$

By hypothesis, $\nabla_{\underline{d}}\langle q(\underline{d}^*)|\underline{d}^*\rangle = t\nabla_{\underline{d}}\langle u(\underline{d}^*)\rangle$ for some positive scalar t; by defining $\underline{\tau} = (1/t)\underline{\lambda}$ and $\underline{\nu} = (1/t)\underline{\mu}$, we have

$$\nabla_{\underline{d}}\langle u(\underline{d}^*)\rangle + \underline{\tau}\nabla_{\underline{d}}\underline{h}(\underline{d}^*) + \underline{\nu}\nabla_{\underline{d}}\underline{g}(\underline{d}^*) = \underline{0},\quad \underline{\nu} \geq \underline{0}.$$

By assumption, D is convex and $\langle u(\underline{d})\rangle$ is concave, so the Second-Order Sufficiency Conditions (Appendix C) hold at $\underline{d}^*$. Therefore $\underline{d}^*$ maximizes $\langle u(\underline{d})\rangle$. Q.E.D.

Theorem 5.1 motivates the proxy iteration algorithm under uncertainty. At each iteration k, we must fit a proxy $\langle q(\underline{d})|\underline{d}^k\rangle$ for the true objective $\langle u(\underline{d})\rangle$. This proxy must generate a feasible direction of improvement in the decision space D; therefore it must satisfy the following property: for some $t > 0$,

$$\nabla_{\underline{d}}\langle q(\underline{d})|\underline{d}^k\rangle = t\nabla_{\underline{d}}\langle u(\underline{d}^k)\rangle.$$

Under uncertainty, the true objective is not a utility function; rather, it is the expected

utility of a lottery as a function of the decision $\underline{d}$. *Consequently, the proxy is not a utility function either; it is an approximation of* $\langle u(\underline{d})\rangle = \int_{\underline{s}} u[\underline{x}(\underline{d},\underline{s})]\ \{\underline{s}|\varepsilon\}$.

At each iteration, we maximize the proxy

$$\max_{\underline{d}\in D} \ \langle [q(\underline{d})|\underline{d}^k]\rangle$$

and continue the procedure until $\underline{d}^{k+1} = \underline{d}^k$.

Unfortunately, we do not know how to construct the proxies $\langle p[\underline{d}|\underline{d}^k]\rangle$ and $\langle q[\underline{d}|\underline{d}^k]\rangle$ to guarantee that

$$\nabla_{\underline{d}}\langle p[\underline{d}|\underline{d}^k]\rangle = t\nabla_{\underline{d}}\langle V(\underline{d}^k)\rangle \tag{5.1}$$

for some $t > 0$, for the expected-value decision maker, and that

$$\nabla_{\underline{d}}\langle q[\underline{d}|\underline{d}^k]\rangle = t\nabla_{\underline{d}}\langle u(\underline{d}^k)\rangle \tag{5.2}$$

for the risk-sensitive decision maker. As a result, we cannot guarantee that each iteration makes an improvement in the true objective $\langle u(\underline{d})\rangle$. In the deterministic problem, the proxy $p(\underline{x})$ always provided a direction of improvement in the outcome space $X(D)$ since $\nabla p(\underline{x}^k)$ was collinear with $\nabla V(\underline{x}^k)$; by construction, $\nabla_{\underline{x}} p(\underline{x}^k) = t\nabla_{\underline{x}} V(\underline{x}^k)$, for some $t > 0$. Under uncertainty, however, the decision maker must choose among lotteries rather than among deterministic outcomes since each decision vector $\underline{d}$ produces the outcome lottery $\{\underline{x}|\underline{d},\varepsilon\}$ as shown in Figure 1.1. Deterministic and risk preferences must be assessed over the outcome variables, not the decision variables. Constructing the proxies to satisfy (5.1) and (5.2) is difficult since $\underline{d}^k$ corresponds not to one outcome $\underline{x}^k$, but to a probability distribution over the outcome space $X(D)$.

A scheme must be devised for choosing the outcome at each iteration where the parameters are to be assessed so the resulting proxies satisfy (5.1) and (5.2). Such a scheme would guarantee a direction of improvement in $\langle u(\underline{d})\rangle$ at each iteration k; it would specify a decision $\underline{d}^{k+1} \in D$ such that $\langle u(\underline{d}^{(k+1)'})\rangle > \langle u(\underline{d}^k)\rangle$ where $\underline{d}^{(k+1)'} = \alpha\underline{d}^k + (1-\alpha)\underline{d}^{k+1}$ for some α, $0 < \alpha \leq 1$. The direction of improvement required in the decision set D does not correspond to any meaningful direction of improvement in the outcome space $X(D)$ since $\underline{d}^k$ and $\underline{d}^{k+1}$ do not correspond to any unique $\underline{x}^k$ and $\underline{x}^{k+1}$.

5.2 A Scheme for Fitting the Proxies

A natural point for assessing the proxy parameters is the conditional expectation $\langle \underline{x}|\underline{d},\varepsilon\rangle$. The following example shows, however, that when fitting the proxies at $\langle \underline{x}|\underline{d},\varepsilon\rangle$, the algorithm may fail to generate a direction of improvement at each iteration. The example uses an expected-value decision maker with one decision variable and one state variable. The decision maker's problem is:

$$\mathrm{Max}_{d\in D} \int_s V[\underline{x}(d,s)] \ \{s|\varepsilon\},$$

where $V(\underline{x})$ is unknown and cannot be assessed. In this example, we use a Cobb-Douglass proxy for a sum-of-exponentials deterministic preference function.

Deterministic preference function: $V(\underline{x}) = -e^{-0.3x_1} - e^{-0.4x_2}$ (5.3)

Probability distribution on s: $\{s|\varepsilon\} = 2s/3$ for $1 \leq s \leq 2$, 0 elsewhere

System Model: $x_1(d,s) = (d-d^2)s$; $x_2(d,s) = ds$

Constraint: $D = \{d \mid 0 < d < 1\}$

The expected value of the true objective, as a function of d, is:

$$\langle V(d)\rangle = (-2/3) \int_1^2 [se^{-0.3(d-d^2)s} + se^{-0.4ds}]ds.$$

After integrating by parts, the definite integral is written as a function of d alone:

$$\Big((-7.4074)/(d-d^2)\Big) \Big([e^{-0.6(d-d^2)}][-0.6(d-d^2) - 1] - [e^{-0.3(d-d^2)}][-0.3(d-d^2) - 1]\Big) + \Big((-4.1667)/d^2\Big) \Big([e^{-0.8d}][-0.8d - 1] - [e^{-0.4d}][-0.4d - 1]\Big) \quad (5.4)$$

Figure 5.1 is a graph of $\langle V(d)\rangle$ for $0.01 \leq d \leq 0.99$.

The Cobb-Douglass proxy, in its additive form, is

$$p[\underline{x}(d,s)] = \beta \ln x_1(d,s) + (1-\beta) \ln x_2(d,s) \quad (5.5)$$

Fitting the proxy at the conditional expectation $\langle \underline{x}|\underline{d},\varepsilon\rangle$,

$$\langle x_1|d,\varepsilon\rangle = \int_s x_1(d,s)\{s|\varepsilon\} = (14/3)(d-d^2) \quad (5.6)$$

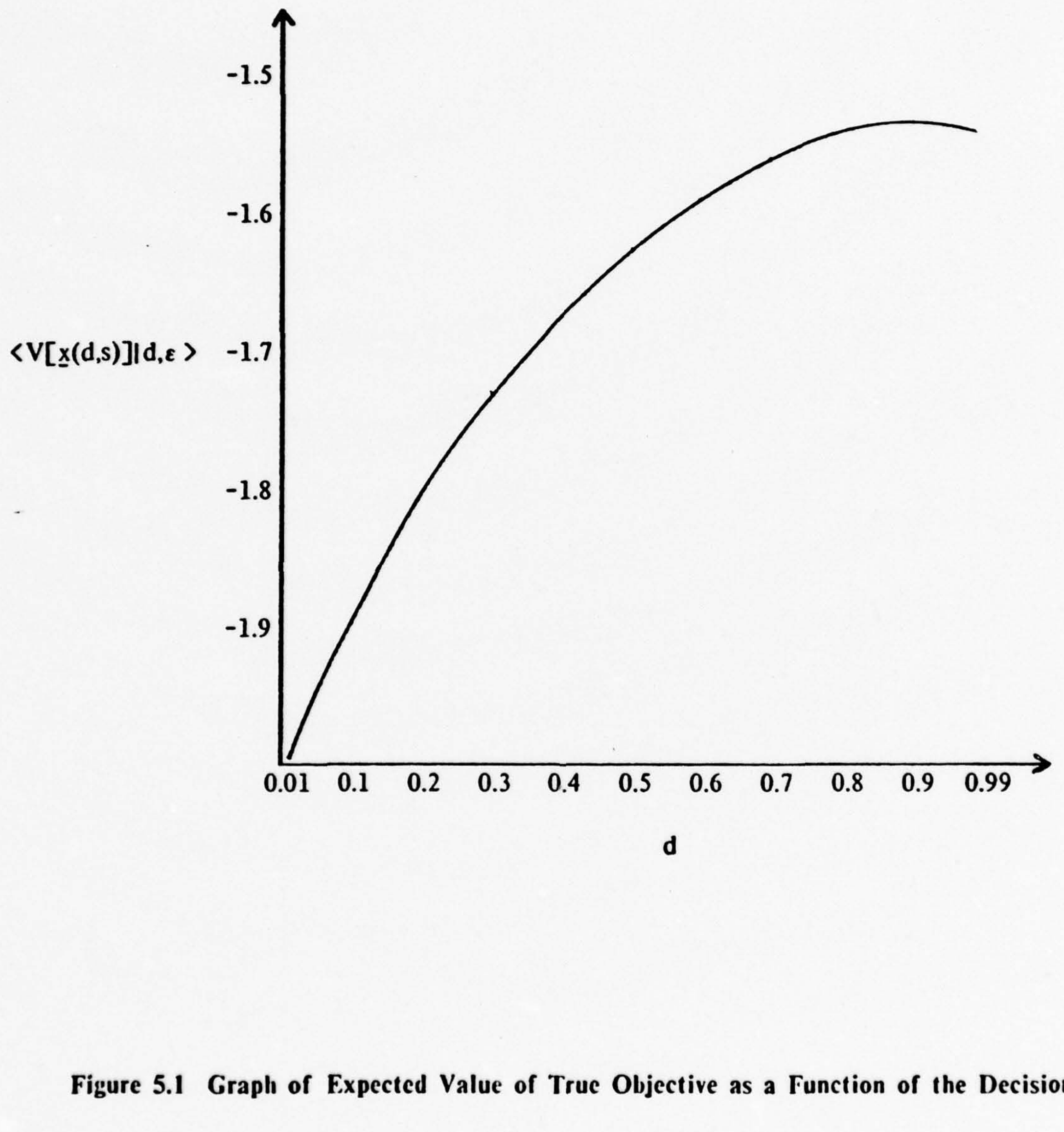


Figure 5.1 Graph of Expected Value of True Objective as a Function of the Decision

$$\langle x_2|d,\varepsilon\rangle = \int_s x_2(d,s)\{s|\varepsilon\} = (14/3)d \tag{5.7}$$

Using the sum-of-exponentials preference function (5.3) to generate tradeoffs at $\langle \underline{x}|d,\varepsilon\rangle$,

$$\lambda_2 = -dx_1/dx_2 = (4/3)\, e^{\,0.3x_1 - 0.4x_2} \tag{5.8}$$

For the Cobb-Douglass proxy (5.5),

$$[\partial p(\underline{x})/\partial x_2]/(\partial p(\underline{x})/\partial x_1] = -dx_1/dx_2 = (1-\beta)x_1/\beta x_2,$$

so

$$\beta = x_1/(x_1 + x_2\lambda_2) \tag{5.9}$$

The expectation of the Cobb-Douglass approximation fit at $\langle \underline{x}|d,\varepsilon\rangle$ is then calculated to find the proxy $\langle p(d)\rangle$ for the true objective $\langle V(d)\rangle$:

$$\langle p(d)\rangle = 0.6667 \int_s [\beta s\cdot \ln(d-d^2)s + (1-\beta)s\cdot \ln(ds)]ds.$$

After integrating by parts, this definite integral is written as a function of d:

$$\Big[0.6667\beta\Big([2\cdot\ln(2d-2d^2) - 1] - [0.5\cdot\ln(d-d^2) - 0.25]\Big) +$$
$$0.6667(1-\beta)\Big([2\cdot\ln(2d) - 1] - [0.5\cdot\ln(d) - 0.25]\Big)\Big]$$

Maximizing $\langle p(d)\rangle$ with respect to d,

$$\partial\langle p(d)\rangle/\partial d = 0 \quad \text{at} \quad d^{\oplus} = 1/(1+\beta). \tag{5.10}$$

For $0 < \beta < 1$, we have $0.5 < d^{\oplus} < 1$, so $d^{\oplus} \in D$.

Equations (5.1) through (5.10) and Figure 5.1 provide the information needed to begin the algorithm.

Choose initial $d^0 = 0.8$.

Fit proxy at $\langle \underline{x}|d^0,\varepsilon\rangle$. Using (5.6) and (5.7),

$$\langle \underline{x}|d^0{=}0.8,\varepsilon\rangle = (0.7467,\ 0.3733).$$

Using (5.8) and (5.9),

$$\lambda_2 = 1.4367 \quad \text{and} \quad \beta = 0.5820,$$

so the proxy for the true objective is

$$\langle p(d)|d^0{=}0.8\rangle \;=\; 0.6667 \int_s [0.582s\cdot\ln(d-d^2)s \;+\; 0.418s\cdot\ln(ds)]ds.$$

Equation (5.10) shows

$$d^1 \;=\; \max_{d\in D} \langle p(d)|d^0{=}0.8\rangle \;=\; 0.632.$$

However,

$$\langle V(d)|d^0{=}0.8\rangle \;=\; -1.5385 \quad\text{and}\quad \langle V(d)|d^1{=}0.632\rangle \;=\; -1.5739,$$

so d^1 is worse than d^0. Figure 5.1 shows the expected value of the true objective decreases monotonically from $d^0 = 0.8$ to $d^1 = 0.632$; no step in this direction, no matter how small, will improve $\langle V(d)\rangle$. Therefore, the algorithm is not globally convergent when the proxy is fit at the conditional expectation $\langle \underline{x}|\underline{d},\varepsilon\rangle$.

Perhaps some other proxy fitting scheme can be found that guarantees improvement at each iteration. Finding such a scheme is a difficult task since the resulting approximation of $u(\underline{x})$ must still be integrated over all outcomes in the lottery $\{\underline{x}|\underline{d},\varepsilon\}$. This task remains the subject of future research.

Boyd [4] tried to use his linear approximation technique for decision making under uncertainty. Instead of fitting the approximation at one point, he tried approximating the risk preference function at all outcomes in the lottery $\{\underline{x}|\underline{d},\varepsilon\}$. At each iteration k, his method requires the assessment of two new parameters,

$$\rho[\underline{x}(\underline{d}^k,\underline{s})] \;=\; [\partial n(\underline{x})/\partial x_1]_{|\underline{x}=\underline{x}(\underline{d},\underline{s})}$$

and

$$\gamma[\underline{x}(\underline{d}^k,\underline{s})] \;=\; [du(n)/dn]_{|n=n[\underline{x}(\underline{d},\underline{s})]}$$

in addition to $\underline{\lambda}$ at *every possible realization* of $\underline{s}$. The resulting assessment requirements are enormous; when $\{s|\varepsilon\}$ is continuous, an infinite number of assessments is required; when $\{\underline{s}|\varepsilon\}$ is discrete, the assessment demands are generally prohibitive since even the simplest binary probability distribution would require six times as many assessments as its deterministic analog (including ρ and γ). Furthermore, the assessment of γ may be a bewildering task for the decision maker.

Consequently, Boyd's procedure is not practical for decision making under uncertainty.

Even if a successful proxy fitting scheme could be found, the analyst must still ask the decision maker at each iteration if the new trial solution is an improvement; at the k^{th} iteration, the decision maker would be asked to choose between the lotteries $\{\underline{x}|\underline{d}^k,\varepsilon\}$ and $\{\underline{x}|\underline{d}^{k+1},\varepsilon\}$. This direct comparison of lotteries requires simultaneous consideration of deterministic tradeoffs and risk attitudes. It is a difficult task for the decision maker to perform, but cannot be avoided since improvement must be guaranteed at each iteration. As a result of these obstacles, the current version of the proxy iteration algorithm is not applicable to problems in which uncertainty plays a major role.

Although the iterative local procedure is not well suited for decision making under uncertainty, other local preference modeling techniques may be useful. Perhaps the analyst can find a "less restrictive" global preference function by assessing local preference models in various decision regions, in the small, and "piecing them together". Some interpolation scheme would be needed to guarantee that the conglomerate global function satisfies the risk preference axioms (2.5-2.9). This research has motivated other work in this direction (in the Stanford Decision Analysis Research Program) and initial results look promising.

CHAPTER VI

PRACTICAL APPLICATION OF THE PROXY ITERATION ALGORITHM

6.1 The Decision Problem

The true practical test of any decision-making procedure is a real problem. In this chapter, I describe the application of the proxy iteration algorithm to a curriculum design problem for the upper grades of a combined elementary and junior high school.

The Shepherd School is a small private school in San Jose, California. It includes a lower school, preschool through fourth grade, and an upper school, fifth grade through eighth grade. Mrs. Wanda Grenke is the head teacher, principal, and administrator at Shepherd School. She is currently planning the curriculum for the upper school and wants to choose the best combination of subjects to teach the students during classroom hours. Her task is a resource allocation problem; how should she allocate the weekly class hours (9:00 am to 3:00 pm, Monday through Friday) to the different subjects: reading, language arts, arithmetic, social studies, science, foreign language, art, music, and physical education. Mrs. Grenke would like to emphasize reading, writing, and arithmetic; even so, she still may choose from an infinite set of possible allocations.

Any allocation she chooses must provide for the following fixed requirements:

1. A fifteen minute recess in the morning and a forty-five minute lunch-recess at midday.
2. One period for chapel every Monday morning, required by the Board of Trustees; the exact length of this period is not specified, but Mrs. Grenke indicates that student concentration seems to dwindle after one hour. She prefers academic periods of 40 to 50 minutes.
3. Three periods of physical education per week (a previously established school policy).
4. One period per week for special projects and class meetings.

Mrs. Grenke also faces constraints on her teaching staff. The Shepherd School, like many other private schools, has a tight operating budget; as a result, the school has a very limited number of teaching positions, some full-time and some part-time. The staff members who teach in the upper school are listed below. For each teacher, the list includes the subjects taught and the number of hours per week he or she can teach.

1. Mrs. Grenke: Reading, language arts, and social studies. Mrs. Grenke is a full-time employee, but her duties as principal and administrator require several hours during each school day. She can devote no more than 16 hours per week to classroom instruction.

2. Mr. Herriman: Mathematics, science, physical education, musical instruments. Mr. Herriman is a full-time staff member. He teaches math, science, and physical education to the upper school, and musical instruments to the lower school. After subtracting time for the fixed physical education and music requirements, he has 19.5 hours remaining during which he can teach math and science in the upper school.

3. Mrs. Findlay: Reading, language arts, social studies. Mrs. Findlay is a part-time staff member working 15 hours per week in the upper school. Her hours are flexible, but must occur in continuous blocks each day without free periods interspersed.

4. Mrs. Williams: Foreign languages. Mrs. Williams is a part-time instructor working in the upper school 3 hours each week. She is available only on Tuesdays and Thursdays midday to early afternoon.

5. Ms. Fyfe: Music. Ms. Fyfe is the full-time kindergarten teacher (lower school) and the music teacher for all grades. The kindergarten closes at lunchtime on Mondays, Wednesdays, and Fridays, so Ms. Fyfe is available these afternoons to teach music to the other grades.

6. Mrs. Blanchard: Art. Mrs. Blanchard is the full-time third-fourth grade teacher and the art teacher for the upper school. She is available to teach art in the upper school on Monday, Wednesday, and Friday afternoons

when Ms. Fyfe instructs her combination third-fourth grade.

No extra funds are available for increasing the staff or the number of hours of the part-time instructors. Mrs. Grenke must design the curriculum given the teaching staff listed above.

The upper school is divided into a fifth-sixth grade combination and a seventh-eighth grade combination. These combined classes are well suited for language arts, social studies, science, foreign language, art, and music since the skills required for these subjects are not grade specific. For reading and arithmetic, however, the combination classes do not work well. The Shepherd School uses the Lippincott Reading Program and Field Mathematics Program; these programs are nationally recognized series of texts for elementary and junior high school instruction. The Lippincott and Field texts are specifically geared for the individual grade levels and cannot be mixed effectively for the combined classes. The Field series comprises Shepherd's entire mathematics program; the Lippincott series comprises about half of Shepherd's reading program. As a result, all mathematics and at least half the reading classes must be taught separately to the fifth, sixth, seventh, and eighth grades.

Mrs. Grenke believes the Shepherd School's most important academic function is the development of its students' verbal and quantitative skills. Consequently, she wants to emphasize the reading and arithmetic programs, but at the same time, she wants to provide a well-balanced education. Just how much of each subject should she include in the upper school curriculum? In section 6.2, I construct a model of her decision problem and in section 6.3, I apply the proxy iteration algorithm to help her find the solution.

6.2 Modeling the Decision Problem

The success or failure of our decision-making procedure (and the global procedure as well) depends on the decision maker's ability to provide tradeoffs that adequately reflect his or her underlying preferences. The decision maker can respond to the assessment questions consistently only if the analyst models the outcomes with objective,

well-defined, and unambiguous attributes. The attributes must be *concrete* enough so the decision maker can easily visualize any particular outcome vector, can clearly distinguish one outcome from another, and can confidently express preferences among outcomes.

In the Shepherd School problem, we must select *concrete* attributes that provide an accurate and complete description of the upper school curriculum. Mrs. Grenke has been a teacher for many years; she has prepared and taught thousands of lesson plans, mostly for forty-five minute class periods. She can clearly visualize a curriculum described by the number of weekly class periods of each subject. Furthermore, she is willing to choose between any two curricula by comparing the amounts of the various subjects. To model her problem, I divide each school day into six class periods, each lasting 45 minutes. I allow three minutes between classes, a fifteen minute recess between periods two and three, and a forty-five minute lunch-recess between periods four and five (in accordance with the fixed requirements enumerated in Section 6.1). These fixed requirements also consume five class periods each week; one for chapel, three for physical education, and one for special topics and class projects. An additional fixed requirement arises from the limited part-time hours of Mrs. Williams, the foreign language instructor. Mrs. Williams can spend only two periods with the fifth-sixth grade and two periods with the seventh-eighth grade each week. She will not work less than three hours (four periods) per week since the resulting compensation would make the arrangement uneconomical. Mrs. Grenke definitely wants to include French in the curriculum, so she must allocate exactly two periods for it each week. These fixed requirements for chapel, physical education, class projects, and foreign language use up seven of the thirty class periods each week. Twenty-three periods remain, to be divided among reading, language arts, arithmetic, social studies, science, art and music. For the purpose of curriculum planning, Mrs. Grenke groups art and music instruction together as a combined program. Consequently, I combine them as one subject in the decision model and represent the upper school curriculum (grades five through eight) by the following six attributes:

x_1 = number of 45 minute periods of reading per week.

x_2 = number of 45 minute periods of language arts per week.

x_3 = number of 45 minute periods of arithmetic per week.

x_4 = number of 45 minute periods of social studies per week.

x_5 = number of 45 minute periods of science per week

x_6 = number of 45 minute periods of art/music each week.

These six attributes, together with the fixed requirements, completely specify the weekly curriculum.

Modeling the Constraints

The fixed requirements consume 7 of the 30 weekly classroom periods, leaving 23 to be allocated to these 6 subjects. This time constraint can be represented quantitatively in terms of the outcome attributes:

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \leq 23.$$

The constraint is written as an inequality since study halls could fill any slack.

The limitations on the teaching staff further constrain the curriculum since certain subjects are taught only by certain teachers. Mrs. Grenke and Mrs. Findlay teach all reading, language arts, and social studies classes in the upper school. Their combined teaching hours are equivalent to 38 forty-five minute periods per week. They can teach language arts, social studies, and half the reading program to the fifth-sixth and seventh-eighth grade combination classes. The other half of the reading program, keyed to the Lippincott readers, requires instruction at four separate levels. The following inequality represents this teaching arrangement:

$$3x_1 + 2x_2 + 2x_4 \leq 38.$$

It implies that the total number of reading, language arts, and social studies classes (at all levels) cannot exceed Mrs. Grenke's and Mrs. Findlay's 38 teaching periods per week. The coefficients indicate the number of levels at which each subject is taught (the coefficient for reading is the average number of levels).

Mr. Herriman handles the entire math and science program of the upper school. The 19.5 hours he has available for these two subjects are equivalent to 24.5 forty-five minute periods (allowing for three minutes between classes). Science lessons are taught to the fifth-sixth and seventh-eighth grade combined classes, but the Field Mathematics Program requires four individual levels. The following inequality constraint indicates that the sum of all math and science periods cannot exceed Mr. Herriman's 24.5 available teaching periods:

$$4x_3 + 2x_5 \leq 24.5.$$

The fixed requirements, for which no tradeoffs are possible, used up 5.5 of his 30 weekly periods.

Mrs. Blanchard and Ms. Fyfe, the art and music instructors, can teach in the upper school during the two afternoon periods three days a week. However, only one of these teachers can teach in the upper school at a time since one must take the third-fourth grade when the other goes to the upper school. Together they can teach the fifth-sixth and seventh-eighth combined grades six times a week. The following inequality indicates that the total number of upper school art and music classes cannot exceed their 6 available periods per week:

$$2x_6 \leq 6.$$

By definition, a nonnegativity constraint holds for each attribute:

$$\underline{x} \geq \underline{0}.$$

Mrs. Grenke indicated that she could eliminate from consideration many allocations that were unquestionably inferior. Without hesitation, she eliminated all curricula that failed to meet the following specifications:

A. Reading, language arts, and arithmetic at least three times a week, but not more than twice a day

B. Social studies at least once a week, but not more than once a day

C. Science not more than once a day

D. Art or music at least once a week, but not more than once a day

Writing these specifications as lower and upper bounds on the subject attributes, we have the following additional constraints:

$$3 \leq x_j \leq 10, \quad j = 1,2,3; \qquad 1 \leq x_k \leq 5, \quad k = 4,6; \qquad 0 \leq x_5 \leq 5$$

Putting all the constraints on class time and teaching staff together with the bounds on the attributes, we define their intersection as the decision set $X(D)$:

$X(D)$ is the set of all $\underline{x}$ such that:

$$\begin{array}{lllllllll} x_1 & + & x_2 & + & x_3 & + & x_4 & + & x_5 & + & x_6 & \leq & 23 \\ 3x_1 & + & 2x_2 & + & & & 2x_4 & & & & & \leq & 38 \\ & & & & 4x_3 & + & & & 2x_5 & & & \leq & 24.5 \end{array}$$

$$\begin{array}{lll} 3 \leq x_1 \leq 10, & 3 \leq x_2 \leq 10, & 3 \leq x_3 \leq 10, \\ 1 \leq x_4 \leq 5, & 0 \leq x_5 \leq 5, & 1 \leq x_6 \leq 3. \end{array}$$

Modeling the objective

Mrs. Grenke's preferences among the possible curricula are transitive and she is willing to make tradeoffs among the six attributes. These tradeoffs are well-defined since each attribute is continuous; any class period can be divided into fractional parts. At any $\underline{x} \in X(D)$, she prefers more of each attribute to less, and her preferences satisfy Axiom 2.4 (decreasing marginal rates of substitution). Thus, her preference ordering over these six attributes restricted to $X(D)$ satisfies the deterministic preference axioms (2.1-2.4) and can therefore be represented by a concave deterministic preference function $V(\underline{x})$.

Mrs. Grenke's decision problem is:

$$\max_{\underline{x} \in X(D)} V(\underline{x})$$

In the next section, I use the proxy iteration algorithm to solve this multi-attribute problem.

Four of the six teachers have constraints not only on the total number of hours

they can teach in the upper school, but also on specific hours certain days of the week. Keeping track of the individual hours of each teacher in the decision model would be a cumbersome task and the resulting optimization algorithm would be horribly complex. Instead, I first solve the allocation problem for $\underline{x} \in X(D)$, and then verify that the solution meets the specific hourly requirements of each teacher. In section 6.4, I develop a class schedule for the upper school that implements the optimal curriculum and satisfies all the time and staff constraints.

6.3 Applying the Proxy Algorithm

Selecting the proxy

Before beginning the iterative procedure, we must select the form of the proxy function that we will use at each iteration. We may choose either the sum-of-exponentials, sum-of-powers, or Cobb-Douglass functions, or even a heterogeneous combination for different attributes. We use the global modeling procedure of Keelin [20] and Barrager [3] to select the functional form that best represents the decision maker's preferences. Our proxy, therefore, has the same form as their global preference function. Using this global procedure, we assume deterministic additivity, assess tradeoffs, and use the estimating formula (B.1) to approximate the marginal value reduction coefficients $z_1, z_2, \ldots, z_6$ (see Appendix B). After assessing z_i at a number of points, we must select one of the following three models of $z_i(x_i)$:

$$z_i(x_i) = \omega_i \quad \Rightarrow \quad v_i(x_i) = -a_i e^{-\omega_i x_i}$$

$$z_i(x_i) = (1 + \alpha_i) / x_i \quad \Rightarrow \quad v_i(x_i) = -a_i x_i^{-\alpha_i}$$

$$z_i(x_i) = 1/x_i \quad \Rightarrow \quad v_i(x_i) = a_i \ln x_i$$

Even if none of the models fits the data, we must still choose one of them since the global procedure can handle only these preference forms. Herein lies a major disadvantage of global preference modeling: the global procedure forces the preferences to fit the function rather than modeling the function to fit the preferences.

For each attribute x_i, we must choose the best model of $z_i(x_i)$. We illustrate the procedure for $z_2(x_2)$, using x_1 as the price variable.

A. Choose a small increment Δx_2 for the tradeoff assessment questions. Ideally, Δx_2 should be the smallest increment over which the decision maker can express meaningful preferences. This optimal increment size is determined experimentally; initially we begin with $\Delta x_2 = 1$.

B. Choose four ordered pairs $(x_1,x_2)^A$, $(x_1,x_2)^B$, $(x_1,x_2)^C$, $(x_1,x_2)^D$ where x_1^A and x_2^A are nominal levels of x_1 and x_2, $x_1^A = x_1^B = x_1^C = x_1^D$, and $x_2^B = x_2^A + \Delta x_2$, $x_2^C = x_2^B + \Delta x_2$, $x_2^D = x_2^C + \Delta x_2$.

C. Assess $\lambda_2(x_1,x_2)$ at pairs A, B, C, and D by finding the Δx_1 at which

$$(x_1 - \Delta x_1, x_2 + \Delta x_2) \sim (x_1, x_2).$$

D. Estimate z_2 using equation (B.1):

$$z_2(x_2)^{AB} \approx (1/\Delta x_2) \ln \{[\lambda_2(x_1,x_2)^A] / [\lambda_2(x_1,x_2)^B\}.$$

Estimate $z_2(x_2)^{BC}$ and $z_2(x_2)^{CD}$ by the same technique.

To estimate z_2, we use an increment of one language arts class period at the following (x_1,x_2) pairs: (7,3), (7,4), (7,5), and (7,6). Table 6.1 shows the tradeoff assessments and the estimates of z_2. We observe that $z_2(3) = 0.51$, $z_2(4) = 0.41$, and $z_2(5) = 0.69$; z_2 is neither a constant, increasing, or decreasing function of x_2. However, the model in which z_i is a constant fits the data better than those in which z_i is decreasing. Therefore we use the exponential form $-a_2e^{-\omega_2 x_2}$ for attribute x_2 in our proxy function.

The same technique was used to estimate z_3. Table 6.1 shows $z_3(4) = 0.69$, $z_3(5) = 0.41$, and $z_3(6) = 0.69$. Again, none of the models provides a very good fit, but the constant parameter form is the best of the three.

For z_4 and z_5, I asked the tradeoff questions using $\Delta x_4 = 0.5$ and $\Delta x_5 = 0.5$. The nominal levels of the attribute pairs were also changed to sample different regions of X(D). Mrs. Grenke had no problem responding to the questions with the smaller increments. Table 6.1 shows $z_4(2.5) = 0.67$, $z_4(3.0) = 1.02$, $z_4(3.5) = 0.81$; $z_5(2.0) =$

Table 6.1 Assessments for Selecting the Proxy Function

	Tradeoff Assessment	Estimate of $z_i(x_i)$
z_2 $\Delta x_2 = 1.0$	$\lambda_2(7.0,3.0) = 1.0$	$z_2(3.0) = 0.51$
	$\lambda_2(7.0,4.0) = 0.6$	$z_2(4.0) = 0.41$
	$\lambda_2(7.0,5.0) = 0.4$	$z_2(5.0) = 0.69$
	$\lambda 2(7.0,6.0) = 0.2$	
z_3 $\Delta x_3 = 1.0$	$\lambda_3(7.0,4.0) = 1.2$	$z_3(4.0) = 0.69$
	$\lambda 3(7.0,5.0) = 0.6$	$z_3(5.0) = 0.41$
	$\lambda_3(7.0,6.0) = 0.4$	$z_3(5.0) = 0.69$
	$\lambda_3(7.0,7.0) = 0.2$	
z_4 $\Delta x_4 = 0.5$	$\lambda_4(6.0,2.5) = 0.7$	$z_4(2.5) = 0.67$
	$\lambda_4(6.0,3.0) = 0.5$	$z_4(3.0) = 1.02$
	$\lambda_4(6.0,3.5) = 0.3$	$z_4(3.5) = 0.81$
	$\lambda_4(6.0,4.0) = 0.2$	
z_5 $\Delta x_5 = 0.5$	$\lambda_5(8.0,2.0) = 0.7$	$z_5(2.0) = 0.68$
	$\lambda_5(8.0,2.5) = 0.5$	$z_5(2.5) = 0.45$
	$\lambda_5(8.0,3.0) = 0.4$	$z_5(3.0) = 0.58$
	$\lambda_5(8.0,3.5) = 0.3$	
z_6 $\Delta x_6 = 0.2$	$\lambda_6(8.0,1.6) = 1.0$	$z_6(1.6) = 0.00$
	$\lambda_6(8.0,1.8) = 1.0$	$z_6(1.8) = 1.12$
	$\lambda_6(8.0,2.0) = 0.8$	$z_6(2.0) = 0.00$
	$\lambda_6(8.0,2.2) = 0.8$	
z_6 $\Delta x_6 = 0.5$	$\lambda_6(8.0,1.5) = 1.3$	$z_6(1.5) = 0.74$
	$\lambda_6(8.0,2.0) = 0.9$	$z_6(2.0) = 0.81$
	$\lambda_6(8.0,2.5) = 0.6$	$z_6(2.5) = 0.81$
	$\lambda_6(8.0,3.0) = 0.4$	
z_1 $\Delta x_1 = 0.5$	$\lambda_1(5.0,6.0) = 1.3$	$z_1(3.0) = 0.52$
	$\lambda_1(5.5,6.0) = 1.0$	$z_1(4.0) = 0.71$
	$\lambda_1(6.0,6.0) = 0.7$	$z_1(5.0) = 0.67$
	$\lambda_1(6.5,6.0) = 0.5$	

0.68, $z_5(2.5) = 0.45$, and $z_5(3.0) = 0.58$. These estimates indicate z_4 and z_5 are certainly not decreasing functions of x_4 and x_5, respectively. Forced to choose between the constant or decreasing z_i models, we select the constant parameter in each case and use the exponential forms $-a_4e^{-\omega_4 x_4}$ and $-a_5e^{-\omega_5 x_5}$ for attributes x_4 and x_5 in the proxy function.

For attribute x_6, I reduced the tradeoff assessment increment to $\Delta x_6 = 0.2$, but the decision maker responded to several questions inconsistently. Table 6.1 shows $\lambda_6(8.0,1.6) = \lambda_6(8.0,1.8)$ and $\lambda_6(8.0,2.0) = \lambda_6(8.0,2.2)$. These tradeoffs yield $z_6(1.6) = 0$, $z_6(1.8) = 1.12$, and $z_6(2.0) = 0$, implying regions of constant marginal rates of substitution. These regions of linear indifference curves violate Axiom 2.4 (decreasing marginal rates of substitution). Apparently, Mrs. Grenke had trouble distinguishing between pairs of attributes whose corresponding components differed by less than 10 minutes.

When the assessment procedure was repeated using $\Delta x_6 = 0.5$, Mrs. Grenke's responses obeyed Axiom 2.4. Table 6.1 shows the new tradeoffs and the resulting values of $z_6(x_6)$: $z_6(1.5) = 0.74$, $z_6(2.0) = 0.81$, and $z_6(2.5) = 0.81$. The constant z_i model gives a good fit for these data points.

Finally, we model z_1 using x_3 as the price variable and $\Delta x_3 = 0.5$ as the assessment increment. Table 6.1 shows the tradeoff assessments and the resulting values of $z_1(x_1)$: $z_1(5.0) = 0.52$, $z_1(5.5) = 0.71$, and $z_1(6.0) = 0.67$. We again find that none of the models is appropriate, but the constant z_i form provides the best fit. None of the z_i parameters show a clearly decreasing pattern, so we use the constant z_i model for each attribute. The resulting proxy is the sum-of-exponentials function:

$$p(\underline{x}) = -\sum_i a_i e^{-\omega_i x_i}.$$

All six attributes are similar entities measured along a common dimension. It is therefore not surprising that the tradeoff structure is similar for each attribute.

Beginning the Iterative Procedure

Having selected the form of the proxy function, we may now begin the first

iteration. We drop the deterministic additivity assumption, so we must specify all six attributes when assessing tradeoffs between any pair.

Mrs. Grenke wants to emphasize reading and arithmetic. She knows the students benefit from the individualized instruction in these two subjects, but wonders whether one reading and one math period per day is sufficient. We use this information to choose the two initial points:

$$\underline{x}^1 = (8,\ 5,\ 6,\ 2,\ 0.5,\ 1.5)$$

$$\underline{x}^2 = (4.5,\ 4,\ 4.5,\ 4,\ 3,\ 3);$$

$\underline{x}^1$ gives a strong emphasis to reading and math, while $\underline{x}^2$ gives them only a slight emphasis relative to the other subjects.

We want the tradeoff assessments to represent the true gradient as accurately as possible. To help Mrs. Grenke conceptualize different outcome vectors, I asked each teacher to provide typical lesson plans for forty-five minute periods in his or her subjects. Mrs. Grenke could refer to these characteristic lesson plans to help visualize fractional parts of periods. In the assessment procedure, the price variable is x_1 and the assessment increment is one-half period. Both the single point and successive point consistency tests are used.

Figure 6.1a shows the tradeoffs assessed at $\underline{x}^1$. To check consistency, a second set of tradeoffs was assessed at $\underline{x}^1$, with x_2 as the price variable. I set an allowable tolerance of 25%; if the discrepancy for any tradeoff exceeds 25%,

$$| \lambda_{1j} - \lambda_{1k} \cdot \lambda_{kj} | \,/\, \lambda_{1j} > 25\%,$$

it must be resolved. These errors are resolved by explaining the inconsistency to the decision maker, showing her the direction in which the violating component must change, and reassessing the tradeoffs until the tolerance condition is satisfied. In cases where λ_{1j} is very small, 0.3 or less, a larger tolerance of 35% is allowed.

Figures 6.1a and 6.1b show the tradeoff assessments and consistency tests at $\underline{x}^1$ and $\underline{x}^2$. At both points, the discrepancies exceeded the 20% tolerance, so several reassessments were required. The attribute x_3 was used as the price variable at $\underline{x}^2$.

$\underline{x}^1$: (8.0, 5.0, 6.0, 2.0, 0.5, 1.5)

Tradeoff assessments at $\underline{x}^1$, with price variable x_1: (1.0, 1.0, 0.6, 1.6, 2.4, 1.4)

Single point consistency check, with price variable x_2

$\lambda_{2j}(\underline{x}^1)$: (0.8, 1.0, 0.8, 1.8, 2.2, 1.0)

Percent error by tradeoff:

20.0 0.0 33.3 12.5 8.3 28.6

Error exceeds tolerance: Reassess violating components

$\lambda_{16}(\underline{x}^1) \leftarrow 1.3$ $\lambda_{23}(\underline{x}^1) \leftarrow 0.6$ $\lambda_{26}(\underline{x}^1) \leftarrow 1.2$

Percent error by tradeoff:

20.0 0.0 0.0 12.5 8.3 7.7

These tradeoffs will be used.

Figure 6.1a First Iteration

$\underline{x}^2$: (4.5, 4.0, 4.5, 4.0, 3.0, 3.0)

Tradeoff assessments at $\underline{x}^2$, with price variable x_1: (1.0, 0.4, 1.0, 0.2, 0.2, 0.2)

Single point consistency check, with price variable x_3

$\lambda_{3j}(\underline{x}^2)$: (0.7, 0.2, 1.0, 0.3, 0.2, 0.2)

Percent error by tradeoff:

30.0 50.0 0.0 50.0 0.0 0.0

Error exceeds tolerance: Reassess violating components

$\lambda_{12}(\underline{x}^2) \leftarrow 0.5$ $\lambda_{13}(\underline{x}^2) \leftarrow 0.9$ $\lambda_{14}(\underline{x}^2) \leftarrow 0.3$ $\lambda_{31}(\underline{x}^2) \leftarrow 1.0$ $\lambda_{32}(\underline{x}^2) \leftarrow 0.6$ $\lambda_{35}(\underline{x}^2) \leftarrow 0.3$

Percent error by tradeoff:

10.0 8.0 0.0 10.0 35.0 10.0

These tradeoffs will be used.

Nearby point $\underline{x}^{2'}$: (6.0, 4.0, 4.5, 4.0, 3.0, 3.0)

Assess single tradeoff at $\underline{x}^{2'}$: $\lambda_{12}(\underline{x}^{2'}) = 0.7$

Fit proxy and perform successive point consistency test:

Proxy is consistent with DMRS

Maximize proxy with convex programming algorithm:

New maximum $\underline{x}^3$ is: 7.58 3.97 5.24 2.84 1.65 1.73

Ask decision maker: $\underline{x}^3 \succ \underline{x}^2$?

Yes

Iteration improved objective. Use new maximum in next iteration.

Figure 6.1b First Iteration

After the single point inconsistencies were resolved, a single tradeoff λ_2 at the nearby point $\underline{x}^{2'}$ = (6, 4, 4.5, 4, 3, 3) was assessed and checked. The successive point consistency check was then used: if the parameters $\underline{a}$ and $\underline{\omega}$ are positive, the proxy is consistent with Axiom 2.4 (decreasing marginal rates of substitution). Figure 6.1 indicates the proxy passed this consistency test.

The convex programming algorithm maximizes the proxy over X(D). Figure 6.1b shows the new trial solution:

$$\underline{x}^3 = (7.58,\ 3.97,\ 5.24,\ 2.84,\ 1.65,\ 1.73)$$

Mrs. Grenke preferred $\underline{x}^3$ to $\underline{x}^2$, so the new maximum was used to begin the next iteration.

Figure 6.2 shows the tradeoff assessments at $\underline{x}^3$ and the consistency tests. The discrepancies again exceeded the allowed tolerance and required further resolution. The proxy fit from these revised tradeoffs passed the successive point consistency test. This new proxy generated the next trial solution $\underline{x}^4$:

$$\underline{x}^4 = (6.87,\ 3.99,\ 5.14,\ 3.00,\ 1.99,\ 2.01).$$

The time constraint and the constraint on Mr. Herriman are active at this point.

When asked to compare $\underline{x}^4$ with $\underline{x}^3$, Mrs. Grenke indicated without hesitation that $\underline{x}^4$ was preferred. This new maximum was used to begin the third iteration.

The tradeoff assessments for this iteration are shown in Figure 6.3. Only one component exceeded the tolerance of the consistency test. After resolving the inconsistency, an additional λ_2 was assessed and checked at $\underline{x}^{4'}$ = (6, 4, 5.13, 3.00, 1.99, 2.01). The proxy fit from these assessments at $\underline{x}^3$, $\underline{x}^4$, and $\underline{x}^{4'}$ was consistent with Axiom 2.4. Figure 6.3 shows the new trial solution $\underline{x}^5$ generated by the third iteration:

$$\underline{x}^5 = (6.66,\ 3.99,\ 5.16,\ 2.84,\ 1.93,\ 2.42).$$

Attributes x_2, x_4, and x_5 changed very little from iteration two to iteration three. Mrs. Grenke preferred $\underline{x}^5$ to $\underline{x}^4$ because the increase in the art/music component more than compensated for the decrease in reading and social studies. Since the new trial solution was an improvement, the next iteration could begin.

$\underline{x}^3$: (7.58, 3.97, 5.24, 2.84, 1.65, 1.73)

Tradeoff assessments at $\underline{x}^3$, with price variable x_1: (1.0, 1.6, 1.2, 1.4, 1.8, 1.8)

Single point consistency check, with price variable x_4

$\lambda_{4j}(\underline{x}^3)$: (0.6, 0.8, 1.0, 1.0, 1.0, 1.4)

Percent error by tradeoff:

16.0 30.0 16.7 0.0 22.2 8.9

Error exceeds tolerance: Reassess violating components

$\lambda_{12}(\underline{x}^3) \leftarrow 1.5$ $\lambda_{13}(\underline{x}^3) \leftarrow 1.1$ $\lambda_{14}(\underline{x}^3) \leftarrow 1.3$ $\lambda_{15}(\underline{x}^3) \leftarrow 1.7$ $\lambda_{16}(\underline{x}^3) \leftarrow 1.6$

$\lambda_{41}(\underline{x}^3) \leftarrow 0.7$ $\lambda_{42}(\underline{x}^3) \leftarrow 1.0$ $\lambda_{43}(\underline{x}^3) \leftarrow 0.9$ $\lambda_{45}(\underline{x}^3) \leftarrow 1.2$

Percent error by tradeoff:

9.0 13.3 6.4 0.0 8.2 13.7

These tradeoffs will be used.

Fit proxy and perform successive point consistency test:

Proxy is consistent with DMRS

Maximize proxy with convex programming algorithm:

New maximum $\underline{x}^4$ is: 6.87 3.99 5.14 3.00 1.99 2.01

Ask decision maker: $\underline{x}^4 > \underline{x}^3$?

Yes

Iteration improved objective. Use new maximum in next iteration.

Figure 6.2 Second Iteration

$\underline{x}^4$: (6.87, 3.99, 5.14, 3.00, 1.99, 2.01)

Tradeoff assessments at $\underline{x}^4$, with price variable x_1: (1.0, 1.2, 1.2, 0.8, 1.0, 1.2)

Single point consistency check, with price variable x_5

$\lambda_{5j}(\underline{x}^4)$: (1.2, 1.0, 1.0, 1.2, 1.0, 1.2)

Percent error by tradeoff:

20.0 16.7 16.7 50.0 0.0 0.0

Error exceeds tolerance: Reassess violating components

$\lambda_{12}(\underline{x}^4) \leftarrow 1.1$ $\lambda_{13}(\underline{x}^4) \leftarrow 1.1$ $\lambda_{14}(\underline{x}^4) \leftarrow 0.9$ $\lambda_{51}(\underline{x}^4) \leftarrow 1.1$ $\lambda_{54}(\underline{x}^4) \leftarrow 1.1$

Percent error by tradeoff:

10.0 9.1 9.1 22.2 0.0 0.0

These tradeoffs will be used.

Nearby point $\underline{x}^{4'}$: (6.00, 4.00, 5.14, 3.00, 1.99, 2.01)

Assess single tradeoff at $\underline{x}^{4'}$: $\lambda_{12}(\underline{x}^{4'}) = 0.8$

Fit proxy and perform successive point consistency test:

Proxy is consistent with DMRS

Maximize proxy with convex programming algorithm:

New maximum $\underline{x}^5$ is: 6.66 3.99 5.16 2.84 1.93 2.42

Ask decision maker: $\underline{x}^5 \succ \underline{x}^4$?

Yes

Iteration improved objective. Use new maximum in next iteration.

Figure 6.3 Third Iteration

Figure 6.4 shows the tradeoff assessments at $\underline{x}^5$. For the first time, the discrepancy with the additional price variable was sufficiently small. However, the successive point consistency test was violated. Figure 6.4 shows several proxy parameters were negative, indicating increasing marginal rates of substitution. The inherent assessment error was too large relative to the distance between $\underline{x}^4$ and $\underline{x}^5$. These two points were too close for Mrs. Grenke to provide mathematically consistent responses. Consequently, we could not fit a sum-of-exponentials proxy at $\underline{x}^4$ and $\underline{x}^5$ that obeyed Axioms 2.1-2.4. Instead, we used the tradeoff vector at $\underline{x}^5$ alone (which passed the single point consistency test) to determine the linear approximation of the indifference curve. With this linear proxy, we took a modified Frank-Wolfe spacer step. The new maximum $\underline{x}^6$, found by a linear programming algorithm, was the extreme point:

$$\underline{x}^6 = (0,\ 19,\ 4,\ 0,\ 0,\ 0).$$

This new maximum was obviously inferior, so we used it only to indicate the direction of search; the Armijo relaxation procedure, with $\alpha = 0.8$, yielded

$$\underline{x}^{6'} = (5.33,\ 6.49,\ 4.93,\ 2.27,\ 1.54,\ 1.94).$$

Mrs. Grenke preferred $\underline{x}^5$ to $\underline{x}^{6'}$, so another relaxation step was taken, yielding

$$\underline{x}^{6''} = (6.39,\ 4.59,\ 5.11,\ 2.73,\ 1.85,\ 2.32).$$

Mrs. Grenke still preferred $\underline{x}^5$ to $\underline{x}^{6''}$, so a third Armijo step was tried, yielding

$$\underline{x}^{6'''} = (6.61,\ 4.11,\ 5.15,\ 2.82,\ 1.91,\ 2.40).$$

Mrs. Grenke could barely distinguish $\underline{x}^{6'''}$ from $\underline{x}^5$; the largest difference in any attribute was just over five minutes. Since the Frank-Wolfe step specifies the best local direction of improvement, and since no distinguishable improvement was found in this direction, we terminated the procedure, declaring $\underline{x}^5$ as the optimal solution.

6.4 Implementing the Optimal Solution

For simplicity, several specific hourly constraints on teachers and subjects were not included in the decision model. Having found the optimal allocation, I tried to design a

$\underline{x}^5$: (6.66, 3.99, 5.16, 2.84, 1.93, 2.42)

Tradeoff assessments at $\underline{x}^5$, with price variable x_1: (1.0, 1.2, 1.0, 0.8, 1.0, 1.0)

Single point consistency check, with price variable x_6

$\lambda_{6j}(\underline{x}^5)$: (1.0, 1.0, 1.2, 1.0, 0.8, 1.0)

Percent error by tradeoff:

0.0 16.7 20.0 25.0 20.0 0.0

These tradeoffs will be used.

Fit proxy and perform successive point consistency test:

Proxy violates DMRS: parameters a_2, a_3, a_6, ω_2, ω_3, ω_6 are negative

Points $\underline{x}^4$ and $\underline{x}^5$ too close

Fit linear proxy at $\underline{x}^5$ and take spacer step

Maximize proxy with linear programming algorithm

New maximum $\underline{x}^6$ is: 0.0 19.0 4.0 0.0 0.0 0.0

Ask decision maker: $\underline{x}^6 > \underline{x}^5$? No

Use relaxation procedure: $\underline{x}^{6'}$: 5.33 6.49 4.93 2.27 1.54 1.94

Ask decision maker if $\underline{x}^{6'} > \underline{x}^5$? No

Use relaxation procedure: $\underline{x}^{6''}$: 6.39 4.59 5.11 2.73 1.85 2.32

Ask decision maker if $\underline{x}^{6''} > \underline{x}^5$? No

Use relaxation procedure: $\underline{x}^{6'''}$: 6.61 4.11 5.15 2.82 1.91 2.40

Ask decision maker if $\underline{x}^{6'''} > \underline{x}^5$? Indifferent

Optimal solution: $\underline{x}^5$

Figure 6.4 Fourth Iteration

class schedule that implemented the optimal curriculum and met all the additional requirements listed below:

A. Chapel must be the first period on Monday.

B. Physical education can only be the last period of the day.

C. Art and music can only be taught during periods five and six on Monday, Wednesday, and Friday afternoons.

D. Foreign language classes must be taught after 11:30 on Tuesdays and Thursdays.

E. Mrs. Findlay's teaching periods must occur in continuous blocks, without interspersed free periods (except lunch).

F. All students must be supervised during the entire day. Therefore, when the combination classes are separated, one grade must have reading while the other has math.

G. *Lunch period should* begin at noon or shortly thereafter (a lower school requirement imposed on the upper school).

H. Five to ten minutes are needed at the beginning of each school day for a homeroom period.

After a considerable amount of juggling teachers, subjects, and hours, I devised an actual class schedule implementing the optimal solution. Figure 6.5 shows this optimal curriculum. The time constraint and the constraints on Mrs. Grenke, Mrs. Findlay, and Mr. Herriman are active, implying all their teaching hours are utilized. The morning periods were shortened slightly to forty-two minutes to provide an earlier lunch break than would have been possible with six uniform forty-five minute periods. The two afternoon periods were lengthened to fifty-four minutes, filling the rest of the day. The longer afternoon periods allow for the optimal art/music allocation without split periods. Only two class periods per week are split between two subjects, but the same instructor teaches both subjects to the same group, so no additional class breaks are required. The weekly class schedule in Figure 6.5 is equivalent to the following allocation expressed in forty-five minute periods:

Fifth-Sixth Grade

	1	2	3	4	5	6
Monday	Chapel	5M 6R	LA	SS	5R 6M	A/M
Tuesday	5R 6M	5M 6R	LA	FL	Sci	PE
Wednesday	5R 6M	5M 6R	LA	SS	5-6 R	PE
Thursday	5R 6M	5M 6R	LA	FL	Sci (32m.) M_{56}(22m.)	PE
Friday	5R 6M	5M 6R	R_{56}(30m.) LA(12m.)	SS	Spec. Proj.	A/M

Seventh-Eighth Grade

	1	2	3	4	5	6
Monday	Chapel	SS	7M 8R	7R 8M	A/M	Sci (32m.) M_{78}(22m.)
Tuesday	LA	7-8 R	7M 8R	7R 8M	FL	PE
Wednesday	LA	SS	7M 8R	7R 8M	Spec. Proj.	PE
Thursday	LA	LA(12m.) R_{78}(30m.)	7M 8R	7R 8M	FL	PE
Friday	LA	SS	7M 8R	7R 8M	A/M	Sci

R: Reading
M: Arithmetic
LA: Language Arts
SS: Social Studies
Sci: Science
FL: Foreign Language
A/M: Art or Music
PE: Physical Education

Homeroom: 9:00 - 9:08
Period 1: 9:11 - 9:53
Period 2: 9:56 - 10:38
Recess: 10:38 - 10:53
Period 3: 10:53 - 11:35
Period 4: 11:38 - 12:20
Lunch-Recess: 12:20 - 1:06
Period 5: 1:09 - 2:03
Period 6: 2:06 - 3:00

Figure 6.5 Implementation of the Optimal Curriculum

$$\underline{x} = (6.65,\ 3.97,\ 5.13,\ 2.79,\ 1.92,\ 2.40).$$

The optimal solution found by the proxy algorithm was:

$$\underline{x}^* = (6.66,\ 3.99,\ 5.16,\ 2.84,\ 1.93,\ 2.42).$$

The actual implementation sums to 22.86 instead of 23 periods; the slight differences in the second decimal result from including the eight minute homeroom period.

Mrs. Grenke was very pleased with the analysis. She believed the solution was truly optimal and implemented it on a regular basis in January, 1977.

CHAPTER VII
SUMMARY

7.1 Conclusions

In this thesis, I have combined two rival preference modeling techniques in a new approach to multi-attribute decision making. This new approach incorporates the normatively motivated preference models of the global procedure as proxy functions in a local procedure. The proxy approach uses the advantages of one technique to overcome the disadvantages of the other; the resulting combined technique yields rapid convergence without restrictive assumptions.

The curriculum planning problem shows the proxy approach is practical for decision making under certainty. The decision maker, previously unfamiliar with decision analysis, was able to provide the assessments required at each iteration. With the help of the consistency tests, the tradeoff assessments generated trial solutions that converged rapidly to the optimum.

7.2 Suggestions for Future Research

In Chapter V, we observed theoretical aspects of the decision problem under uncertainty that present major obstacles for the proxy algorithm. Although the iterative local procedure is not well suited for decision making under uncertainty, other local preference modeling procedures may be useful. Perhaps the analyst could construct local preference models in different regions of the outcome space and piece them together to form a less restrictive global preference function. Design of an interpolation scheme guaranteeing the conglomerate global function satisfies the risk preference axioms would be an important practical contribution.

A second area deserving more attention is the selection of the outcome attributes themselves. The proxy iteration algorithm takes the outcome variables as given, and addresses the problem of finding the optimal decision; it does not focus on construction of the decision model itself. Selecting the decision variables, state variables, and

outcome variables, and modeling the relationships among them are crucial parts of a decision analysis. Various hierarchical and economic modeling schemes have been tried, but no unified approach to modeling multi-attribute problems currently exists. Building a general framework to help identify the key elements of complex decisions remains an important topic for future research.

Appendix A

NOTATION

x_i	one-dimensional outcome variable
$\underline{x}$	multi-attribute outcome vector; the underscore indicates column vector
$\underline{x}^T$	transpose of $\underline{x}$
$\underline{d}$	vector of decision variables
D	set of feasible decisions $\underline{d}$
$X(D)$	subset of Euclidean space restricted to D
$\underline{s}$	vector of state variables
$\{\underline{s}\|\varepsilon\}$	joint probability distribution over $\underline{s}$ given state of information ε
$\langle\underline{x}\|\varepsilon\rangle$	expectation of $\underline{x}$ given ε
$\langle\underline{x}\|\underline{d},\varepsilon\rangle$	conditional expectation of $\underline{x}$ given $\underline{d}$ and ε
$\lambda_{ij}(\underline{x})$	marginal rate of substitution of x_i for x_j at $\underline{x}$
$\lambda_j(\underline{x})$	marginal rate of substitution of x_1 for x_j at $\underline{x}$
$V(\underline{x})$	deterministic preference function
$n(\underline{x})$	numeraire function
$u(\underline{x})$	risk preference function
$h[\underline{x} \mid \underline{x}^k]$	linear approximation of deterministic preference function fit at $\underline{x}^k$
$p[\underline{x} \mid \underline{x}^k,\underline{x}^{k-1}]$	sum-of-exponentials approximation of deterministic preference function, fit from tradeoff assessments at $\underline{x}^k$ and $\underline{x}^{k-1}$
$q[\underline{d}\|\underline{d}^k,\underline{s}]$	approximation of risk preference function fit at $\underline{x}(\underline{d}^k,\underline{s})$
$\succ$	strict preference relation
$\sim$	indifference relation
$\succsim$	weak preference relation
$>$	is greater than
$=$	is equal to
$\geq$	is greater than or equal to

Appendix A: Notation

$\{\underline{x}^k\}$	infinite sequence in $\underline{x}$
$\underline{x}^\dagger$	limit point of $\{\underline{x}^k\}$
∇f	gradient of the function f, defined as a row vector
$\nabla^2 f$	Hessian matrix of f
$\int$	general summation operator
$\int f(x)$	indefinite integral of $f(x)$
$\sum_i f_i$	summation of f_i over i; *lower limit is* 1 *and upper limit is* N *unless specified otherwise*
$\equiv$	is identically equal to

APPENDIX B

BEHAVIORAL PROPERTIES OF SEVERAL PREFERENCE FUNCTIONS

The Sum-of-Exponentials, Sum-of-Powers, and Cobb-Douglass Deterministic Preference Functions

With Axiom 2.4, we assumed decreasing marginal rates of substitution: as the individual accumulates increasing amounts of attribute x_j, the marginal price dx_i he is willing to pay for each additional dx_j declines. This property alone does not limit the preference function to a specific form. A measure of the rate at which the marginal rate of substitution λ_{ij} decreases is needed. Keelin's marginal value reduction coefficient [20],

$$z_{ij}(\underline{x}) = -[\partial\lambda_{ij}(\underline{x})/\partial x_j] \;/\; \lambda_{ij}(\underline{x}),$$

measures this property; it is defined as the percentage decrease in $\lambda_{ij}(\underline{x})$ per unit increase in x_j with all other attributes held constant.

Keelin and Barrager each assume deterministic independence. For $N \geq 3$, this very strong assumption states that preferences among any subset of attributes are independent of fixed levels of the other attributes. This assumption guarantees the preference function has an additive form, $V(\underline{x}) = \Sigma_i \; v_i(x_i)$.

Combining various assumptions about the coefficient $z_{ij}(\underline{x})$ with deterministic additivity, Keelin and Barrager derive the sum-of-exponentials, sum-of-powers and Cobb-Douglass preference functions. The assumptions are stated below for each function; the derivations are found in their dissertations [3],[20].

Sum-of-Exponentials. If $z_{ij}(\underline{x})$ is a positive constant ω_j for all $j \neq i$, and if $V(\underline{x}) = \Sigma_i \; v_i(x_i)$, then $V(\underline{x}) = -\Sigma_i \; a_i e^{-\omega_i x_i}$.

For each additional Δx_j, the percentage decrease in λ_{ij} is constant at all $\underline{x}$, independent of the level of any attribute. This preference function obeys the following property:

$$\underline{x}^1 \; I \; \underline{x}^2 \quad \Rightarrow \quad \underline{x}^1 + \Delta' \quad I \quad \underline{x}^2 + \Delta'$$

where $\Delta' = (\Delta, \Delta, \ldots, \Delta)$.

Its indifference curves are convex if and only if $\underline{\omega}, \underline{a} > \underline{0}$.

Proof: Along any indifference curve, $\partial[dx_i/dx_j] / \partial x_j = [(\omega_j^2 a_j)/(\omega_i a_i)]\, e^{-\omega_j x_j} e^{\omega_i x_i}$. If $\underline{\omega}, \underline{a} > \underline{0}$, each term is positive, so $\partial[dx_i/dx_j] / \partial x_j > 0$ for all i,j. For the converse, Axiom 2.4 implies $\partial[dx_i/dx_j] / \partial x_j > 0$ for all i,j. Nonsatiety implies $\partial V(\underline{x})/\partial x_j > 0$, so $\omega_j a_j > 0$ for all j. Therefore $[(\omega_j^2 a_j)/(\omega_i a_i)]\, e^{-\omega_j x_j} e^{\omega_i x_i} > 0 \;\Rightarrow\; \omega_j, a_j > 0 \;\Rightarrow\; \underline{\omega}, \underline{a} > \underline{0}$. Q.E.D.

Keelin's exponential estimate of $z_{ij}(\underline{x})$:

For $v_j(x_j) = -a_j e^{-\omega_j x_j}$, the marginal value reduction coefficient is $z_{ij}(\underline{x}) = z_j(x_j) = \omega_j$. Let $\underline{x} = (x_1, x_2, \ldots, x_j, \ldots, x_N)$ and $\underline{x}^{\oplus} = (x_1, x_2, \ldots, x_{j-1}, x_j + \Delta x_j, x_{j+1}, \ldots, x_N)$. Taking the ratio of λ_{ij} at these points,

$$\lambda_{ij}(\underline{x}) \;/\; \lambda_{ij}(\underline{x}^{\oplus}) \quad = \quad e^{\omega_j \Delta x_j},$$

so

$$z_{ij}(\underline{x}) \quad = \quad \omega_j \quad = \quad (1/\Delta x_j) \ln [\lambda_{ij}(\underline{x}) \,/\, \lambda_{ij}(\underline{x}^{\oplus})]. \tag{B.1}$$

This equation provides an exponential estimate for $z_{ij}(\underline{x})$.

Sum-of-Powers: If $z_{ij}(\underline{x}) = (1+\alpha_j)/x_j$ for all $j \neq i$, and if $V(\underline{x}) = \Sigma_i\, v_i(x_i)$, then $V(\underline{x}) = -\Sigma_i\, a_i x_i^{-\alpha_i}$ when $\alpha_j \neq 0$, and $V(\underline{x}) = \Sigma_i\, a_i \ln x_i$ when $\alpha_j = 0$.

With these functions, an additional unit of any attribute implies a smaller percentage decrease in its marginal value. As the individual accumulates more of each attribute, he becomes less sensitive to substitutions among them.

For the special case when $\alpha_j = 0$, the preference function $V(\underline{x}) = \Sigma_i\, a_i \ln x_i$ is the additive form of the **Cobb-Douglass** function $V(\underline{x}) = \Pi_i\, x_i^{a_i}$. Both functions define the same preference ordering since they are equivalent via a logarithmic transformation. For the Cobb-Douglass function, preferences are invariant under scaling: if $\underline{x}^1 \; I \; \underline{x}^2$, then $b\underline{x}^1 \; I \; b\underline{x}^2$ for any positive constant b. This invariance also holds for scaling on a single

attribute.

Each of these preference functions is strictly concave since its Hessian matrix is negative definite. The indifference surfaces associated with each preference function are strictly convex since the marginal rates of substitution are strictly decreasing.

State-of-the-art procedures discussed in section 2.2 use these functions as global preference models. The underlying assumptions are very restrictive in the large, but they are quite reasonable and provide good proxy functions in the small.

The Keeney-Raiffa-Fishburn Independence Conditions [11],[20],[21]

Definition. Risk Additive Independence: Attributes $x_1, x_2, \ldots, x_N$ are risk additive independent if preferences over lotteries on $\underline{x}$ depend only on the marginal probability distributions $\{x_1|\varepsilon\}, \{x_2|\varepsilon\}, \ldots, \{x_N|\varepsilon\}$ and not on the joint probability distribution $\{x_1, x_2, \ldots, x_N|\varepsilon\}$.

Theorem. A multi-attribute utility function is additive, $u(\underline{x}) = \Sigma_i\, u_i(x_i)$, if and only if the attributes $x_1, x_2, \ldots, x_N$ are risk additive independent.

Definition. Utility Independence: Let Y be a subset of $x_1, x_2, \ldots, x_N$. Then Y is utility independent of its complement Y^c if for any pair of lotteries $\{Y|\varepsilon_1\}$ and $\{Y|\varepsilon_2\}$,

$$\{Y|\varepsilon_1\} > \{Y|\varepsilon_2\} \text{ for some fixed } Y^c \quad \Rightarrow \quad \{Y|\varepsilon_1\} > \{Y|\varepsilon_2\} \text{ for any } Y^c$$

where $>$ indicates strict preference over lotteries. Attributes $x_1, x_2, \ldots, x_N$ are **mutually utility independent** if every subset of $x_1, x_2, \ldots, x_N$ is utility independent of its complement.

Theorem. A multi-attribute utililty function has either an additive or multiplicative form, $u(\underline{x}) = \Sigma_i\, u_i(x_i)$ or $u(\underline{x}) = \Pi_i\, u_i(x_i)$, if and only if the attributes $x_1, x_2, \ldots, x_N$ are mutually utility independent.

APPENDIX C

GENERAL OPTIMIZATION THEOREMS USED IN THIS THESIS

Necessary and Sufficient Conditions for Optimality

Definition. Let $\underline{x}^* \in E^n$ be a point satisfying the constraints

$$\underline{h}(\underline{x}^*) = \underline{0}, \quad \underline{g}(\underline{x}^*) \leq \underline{0},$$

where $\underline{h}(\underline{x})$ represents a set of m equalities $h_1(\underline{x}) = 0$, $h_2(\underline{x}) = 0$, ..., $h_m(\underline{x}) = 0$, and $\underline{g}(\underline{x})$ represents a set of p inequalities $g_1(\underline{x}) \leq 0$, $g_2(\underline{x}) \leq 0$, ..., $g_p(\underline{x}) \leq 0$. Let J be the set of indices j for which $g_j(\underline{x}^*) = 0$. Then $\underline{x}^*$ is said to be a **regular point** of the constraints if the gradient vectors $\nabla h_i(\underline{x}^*)$, $\nabla g_j(\underline{x}^*)$, $1 \leq i \leq m$, $j \in J$ are linearly independent.

Kuhn-Tucker Conditions. Let $\underline{x}^*$ be a relative minimum point for the problem

$$\text{minimize} \quad f(\underline{x})$$
$$\text{subject to} \quad \underline{h}(\underline{x}) = \underline{0}, \quad \underline{g}(\underline{x}) \leq \underline{0}$$

and suppose $\underline{x}^*$ is a regular point of the constraints. Then there is a vector $\underline{\lambda} \in E_m$ and a vector $\underline{\mu} \in E_p$ with $\underline{\mu} \geq \underline{0}$ such that

$$\nabla f(\underline{x}^*) + \underline{\lambda} \nabla \underline{h}(\underline{x}^*) + \underline{\mu} \nabla \underline{g}(\underline{x}^*) = \underline{0}$$

$$\underline{u}\underline{g}(\underline{x}^*) = 0.$$

Second-Order Necessary Conditions. Suppose that the functions $f, g, \underline{h} \in C^2$ and that $\underline{x}^*$ is a regular point of the constraints. If $\underline{x}^*$ is a relative minimum point, then there is a $\underline{\lambda} \in E_m$, $\underline{\mu} \in E_p$, $\underline{\mu} \geq 0$ such that the Kuhn-Tucker conditions hold and such that

$$L(\underline{x}^*) = F(\underline{x}^*) + \underline{\lambda} H(\underline{x}^*) + \underline{\mu} G(\underline{x}^*)$$

is positive semidefinite on the tangent subspace of the active constraints at $\underline{x}^*$.

Second-Order Sufficiency Conditions. Let $f, g, \underline{h} \in C^2$. Sufficient conditions that a point $\underline{x}^*$ be a strict relative mimimum point are that $\underline{x}^*$ be a regular point of the

constraints, that the Kuhn-Tucker conditions hold, and that the Hessian of the Lagrangian $L(x^*)$ is positive definite on the subspace

$$M' = \{\underline{y} \mid \nabla \underline{h}(\underline{x}^*)\underline{y} = \underline{0},\ \nabla g_j(\underline{x}^*)\underline{y} = 0 \quad \text{for all} \quad j \in J\}$$

where

$$J = \{j \mid g_j(\underline{x}^*) = 0,\ u_j > 0\}.$$

Optimization by Dual Methods

Convex Duality Theorem: Suppose the problem

$$\min f(\underline{x})$$

$$\text{subject to} \quad \underline{h}(\underline{x}) = \underline{0}, \quad \underline{g}(\underline{x}) \leq \underline{0}$$

has a local solution at $\underline{x}^*$ with corresponding value r^* and Lagrange multipliers $\underline{\lambda}^*$ and $\underline{\mu}^* \geq \underline{0}$. Suppose also that $\underline{x}^*$ is a regular point of the constraints and the Hessian of the Lagrangian $L(\underline{x}) = F(\underline{x}) + \underline{\lambda}H(\underline{x}) + \underline{\mu}G(\underline{x})$ is positive definite everywhere. Then $\underline{x}^*$ is a global minimum, and the dual problem

$$\text{maximize} \quad \varphi(\underline{\lambda}, \underline{\mu}), \quad \underline{\mu} \geq \underline{0}$$

where

$$\varphi(\underline{\lambda}, \underline{\mu}) = \min_{\underline{x}} \ [f(\underline{x}) + \underline{\lambda}\underline{h}(\underline{x}) + \underline{\mu}\underline{g}(\underline{x})]$$

has a global solution at $\underline{\lambda}^*$, $\underline{\mu}^*$ with corresponding value r^* and $\underline{x}^*$ as the point corresponding to $\underline{\lambda}^*$, $\underline{\mu}^*$ in the definition of φ.

Global Convergence

Global Convergence Theorem: Let A be an algorithm on X, and suppose that given $\underline{x}^0$ the sequence $\{\underline{x}^k\}$ is generated satisfying

$$\underline{x}^{k+1} \in A(\underline{x}^k).$$

Let a solution set $\Gamma \subset X$ be given, and suppose

i. all points $\underline{x}^k$ are contained in a compact set $S \subset X$.

ii. there is a continuous function Z on X such that

a. if $\underline{x} \notin \Gamma$, then $Z(\underline{y}) < Z(\underline{x})$ for all $\underline{y} \in A(\underline{x})$.

b. if $\underline{x} \in \Gamma$, then $Z(\underline{y}) \leq Z(\underline{x})$ for all $\underline{y} \in A(\underline{x})$.

iii. the mapping A is closed at points outside Γ.

Then the limit of any convergent subsequence of $\{\underline{x}^k\}$ is a solution.

Spacer Steps

Spacer Step Theorem: Suppose B is an algorithm on X which is closed outside the solution set Γ. Let Z be a descent function corresponding to B and Γ. Suppose that the sequence $\{x^k\}$ is generated satisfying

$$\underline{x}^{k+1} \in B(\underline{x}^k)$$

for k in an infinite index set K, and that

$$Z(\underline{x}^{k+1}) \leq Z(\underline{x}^k)$$

for all k. Suppose also that the set $S = \{\underline{x} \mid Z(\underline{x}) \leq Z(\underline{x}^0)\}$ is compact. Then the limit of any convergent subsequence of $\{\underline{x}^k\}^K$ is a solution.

Note: All theorems in Appendix C are taken from Luenberger [26].

APPENDIX D

GLOBAL CONVERGENCE OF THE GOLDSTEIN AND ARMIJO PROCEDURES

Consider the optimization problem

$$\text{minimize}_{\underline{x} \in X} \ f(\underline{x}),$$

$$X = \{\underline{x} \mid \underline{x} \in R^n, \ \underline{g}(\underline{x}) \leq \underline{0}\} \qquad \text{(D.1)}$$

where $f:R^n \rightarrow R$ and $g:R^n \rightarrow R^m$. Assume the functions f and g are differentiable, and $X \subset R^n$ is nonempty.

The **Frank-Wolfe** algorithm generates at each iteration a direction of search $\underline{d}^i$ with the following properties:

a. $\underline{d}^i \in R^n$ is bounded, i.e. $\|\underline{d}^i\| \leq B$ for some $B > 0$

b. $\exists \mu^{\diamond} > 0 \ni \forall i, \ \underline{x}^i + \mu \underline{d}^i \in X \ \ \forall \mu \in [0, \mu^{\diamond}]$

c. $\nabla f(\underline{x}^i)\underline{d}^i < 0$

Take any $b \in (0,1/2)$ and define the following function $h:R \rightarrow R$

$$h^i(\lambda) = [f(\underline{x}^i + \lambda\mu^i\underline{d}^i) - f(\underline{x}^i)] \ / \ [\lambda\mu^i \nabla f(\underline{x}^i)\underline{d}^i]$$

where $\mu^i \in [\mu^{\diamond}/\rho, \mu^{\diamond}]$, with $\mu^{\diamond}$ defined above and ρ any given number bigger than one.

Armijo procedure:

For some given $\gamma > 1$ pick $\lambda_i = \max \{1, 1/\gamma, ..., 1/\gamma^n, ...\}$ such that $h^i(\lambda^i) \geq b$ and set $\underline{x}^{i+1} = \underline{x}^i + \lambda^i\mu^i\underline{d}^i$.

Goldstein procedure:

If $h^i(1) \geq b$, pick $\lambda^i = 1$. Otherwise pick any $\lambda \in \Lambda = \{\lambda \mid \lambda \in (0,1), 1-b \geq h^i(\lambda) \geq b\}$.

Set $\underline{x}^{i+1} = \underline{x}^i + \lambda^i\mu^i\underline{d}^i$.

Appendix D: Convergence

Theorem: Convergence of the Armijo Procedure

Let $f{:}R^n \rightarrow R$ be differentiable and continuous on an open set Γ containing X, the constraint set of problem (D.1). Then for any accumulation point $(\underline{x}^\dagger,\underline{d}^\dagger)$ of the sequence $\{\underline{x}^i,\underline{d}^i\}$, $\nabla f(\underline{x}^\dagger)\underline{d}^\dagger = 0$ for the Armijo procedure given above.

Theorem. Convergence of the Goldstein Procedure

Let $f{:}R^n \rightarrow R$ be differentiable and continuous on an open set Γ containing X, the constraint set of problem (D.1). Then for any accumulation point $(\underline{x}^\dagger,\underline{d}^\dagger)$ of the sequence $\{\underline{x}^i,\underline{d}^i\}$, $\nabla f(\underline{x}^\dagger)\underline{d}^\dagger = 0$ for the Goldstein procedure above.

The definitions and theorems in this appendix are taken from Garcia-Palomares [13]. He proves global convergence by the Topkis-Vienott approach [32], showing every accumulation point is a solution.

APPENDIX E

DYER'S EXTENSION OF THE FRANK-WOLFE ALGORITHM

a. Let $\underline{z}^k = \alpha^k \nabla f(\underline{x}^k)^T + \underline{\eta}^k$ be substituted for $\nabla f(\underline{x}^k)$ in the modified Frank-Wolfe algorithm (MFW).

b. Let $\Gamma(\delta) = \{\underline{x} \in X \mid f(\underline{x}) > f(\underline{y}) - \delta \ \ \forall \underline{y} \in X\}$

c. Let $c(\underline{y}) = \max_{\underline{y} \in X} \ \alpha^k \nabla f(\underline{x}^k)\underline{y} + \underline{\eta}^k \cdot \underline{y}$

d. Dyer [9] shows $\lim_{k \to \infty} c(\underline{y}^k) = c(\underline{y}^\infty)$ and proves the following theorem.

Theorem. *If f is concave and differentiable, X is compact and convex,* $\lim_{k\to\infty} \underline{\eta}^k = \underline{\eta}^\infty$, $\lim_{k\to\infty} \alpha^k = \alpha^\infty$, and $\|\underline{\eta}^\infty\| < \infty$, $0 < \|\alpha^\infty\| < \infty$, then MFW either terminates at some finite iteration k and $\underline{x}^k \in \Gamma[(\xi^k/\alpha^k)\|\underline{\eta}^k\|]$, where $\xi^k = \max_{\underline{x} \in X} \|\underline{x} - \underline{y}^k\|$, or MFW generates an infinite sequence $\{\underline{x}^k\}$, and every accumulation point $\underline{x}^\infty$ of $\{\underline{x}^k\}$ is contained in $\Gamma[(\xi^\infty/\alpha^\infty)\|\underline{\eta}^\infty\|]$, where $\xi^\infty = \max_{\underline{x} \in X} \|\underline{x} - \underline{y}^\infty\|$, and $\underline{y}^\infty$ is defined in (d).

REFERENCES

[1] Amor, Jean-Pierre, "On the Initial Rate of Convergence of Large-Step Methods of Feasible Directions," Discussion Paper No. 17, Operations Research Study Center, Graduate School of Management, University of California, Los Angeles, October 1971.

[2] Arrow, Kenneth, and Hurwicz, Leonid, "Decentralization and Computation in Resource Allocation," in R. W. Pfouts (ed.), *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*, Chapel Hill, University of North Carolina Press, 1960, pp. 34-104.

[3] Barrager, Stephen M., *Preferences for Dynamic Lotteries: Assessment and Sensitivity*, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, August 1975.

[4] Boyd, Dean W., *A Methodology for Analyzing Decision Problems Involving Complex Preference Assessments*, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, May 1970.

[5] Brown, Rex V., Kahr, Andrew S., and Peterson, Cameron, *Decision Analysis for the Manager*, New York, Holt, Rinehart and Winston, 1974.

[6] Cazalet, Edward G., *Decomposition of Complex Decision Problems with Applications to Electrical Power Systems Planning*, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, May 1970.

[7] Debreu, Gerard, "Representation of a Preference Ordering by a Numerical Function," in R. M. Thrall (ed.), *Decision Processes*, New York, John Wiley and Sons, 1954, pp. 159-165.

[8] Dyer, James S., "An Empirical Investigation of a Man-Machine Interactive Approach to the Solution of the Multiple-Criteria Problem," Discussion Paper, Western Management Science Institute, University of California, Los Angeles, October 1972.

[9] Dyer, James S., "The Effects of Errors in the Estimation of the Gradient on the Frank-Wolfe Algorithm, with Implications for Interactive Programming," *Operations Research*, Vol. 22, No. 1, Jan-Feb. 1974, pp. 160-174.

[10] Everett, Hugh III, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, Vol. 11, No. 3, May-June 1963, pp. 101-111.

[11] Fishburn, Peter C., and Keeney, Ralph L., "Seven Independence Concepts and Continuous Multi-attribute Utility Functions," *Journal of Mathematical Psychology*, Vol. 11, 1974, pp. 294-327.

[12] Frank, Marguerite, and Wolfe, Philip, "An Algorithm for Quadratic Programming," *Naval Research Logistics Quarterly*, Vol. 3, 1956, pp. 95-110.

[13] Garcia-Palomares, Ubaldo M., *Superlinearly Convergent Quasi-Newton Methods for Nonlinear Programming*, Ph.D. Dissertation, Department of Computer Sciences, University of Wisconsin, 1973.

[14] Geoffrion, Arthur M., Dyer, James S., and Feinberg, A., "An Interactive Approach for Multi-Criterion Optimization, with an Application to the Operation of an Academic Department," *Management Science*, Vol. 19, No. 4, December 1972, pp. 357-368.

[15] Hogan, William W., "Convergence Results for Some Extensions of the Frank-Wolfe Method," Working Paper No. 169, Western Management Science Institute, University of California, Los Angeles, January 1971.

[16] Howard, Ronald A., "The Foundations of Decision Analysis," *IEEE Transactions on Systems Science and Cybernetics*, Vol. SSC-4, No. 3, September 1968, pp. 211-219.

[17] Howard, Ronald A., and Matheson, James E. (eds.), *Readings in Decision Analysis*, Stanford Research Institute, Menlo Park, 1976.

[18] Howard, Ronald A., "Risk Preference," in *Readings in Decision Analysis*, Stanford Research Institute, Menlo Park, 1976, pp. 321-373.

[19] Johnson, Richard, "Trade-Off Analysis: A Method for Quantifying Consumer Values," Market Facts, Inc., 1973.

[20] Keelin, Thomas W., *A Protocol and Procedure for Assessing Multi-Attribute Preference Functions*, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, September 1976.

[21] Keeney, Ralph L., and Raiffa, Howard, *Decision Analysis with Multiple Conflicting Objectives: Preferences and Value Tradeoffs*, WP-75-53, International Institute for Applied Systems Analysis, Laxenburg, Austria, May 1975.

[22] Kruskal, Joseph B., "Nonmetric Multidimensional Scaling: A Numerical Method," *Psychometrika*, Vol. 29, No. 2, June 1964, pp. 1-27.

[23] Luce, R. Duncan, and Suppes, Patrick, "Preferences, Utility, and Subjective Probability," in R. D. Luce, R. R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology*, Vol. III, New York, John Wiley and Sons, Inc., 1965, pp. 249-410.

[24] Luce, R. Duncan, and Tukey, John W., "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement," *Journal of Mathematical Psychology*, Vol. 1, 1964, pp. 1-27.

[25] Luenberger, David G., *Optimization by Vector Space Methods*, New York, John Wiley and Sons, Inc., 1969.

[26] Luenberger, David G., *Introduction to Linear and Nonlinear Programming*, Reading, Mass., Addison-Wesley, 1973.

[27] Morgenstern, Oscar, and von Neumann, John, *Theory of Games and Economic Behavior*, 2nd Ed., Princeton University Press, 1947.

[28] Pratt, John W., "Risk Aversion in the Small and in the Large," *Econometrica*, Vol. 32, No. 1-2, January-April 1964, pp. 122-136.

[29] Raiffa, Howard, "Preferences for Multi-Attributed Alternatives," Rand Corporation, Memorandum RM-5868-DOT/RC, April 1969.

[30] Roy, B., "Problems and Methods with Multiple Objective Functions," *Mathematical Programming*, Vol. 1, No. 2, November 1971, pp. 239-266.

[31] Savage, Leonard, *The Foundations of Statistics*, New York, John Wiley and Sons, 1964.

[32] Topkis, Donald M., and Vienott, Arthur, "On the Convergence of Some Feasible Directions Methods for Nonlinear Programming," *SIAM Journal of Control*, Vol. 5, 1967, pp. 268-279.

[33] Wehrung, Donald A., *Mathematical Programming Procedures for the Interactive Identification and Optimization of Preferences in a Multi-Attributed Decision Problem*, Ph.D. Dissertation, Graduate School of Business, Stanford University, May 1975.

[34] Wolfe, Philip, "Convergence Theory in Nonlinear Programming," Chapter 1 in J. Abadie (ed.), *Integer and Nonlinear Programming*, American Elsevier Publishing Co., New York, 1970.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D. C. 20375	6 copies
Office of Naval Research Attention: Code 102IP 800 North Quincy Street Arlington, Virginia 22217	6 copies

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Assistant Director (Environmental and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: COL Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce MacDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Department of the Army

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson (PERI-OK)
1300 Wilson Boulevard
Arlington, VA 22209

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Other Institutions

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Professor Richard Meyer
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: EES DA-77-3	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): "A Proxy Approach to Multi-Attribute Decision Making"	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: EES DA-77-3
7. AUTHOR(s): Kenneth R. Oppenheimer	8. CONTRACT OR GRANT NUMBER(s): 75-030-0713 N00014-76-C-0074 PER FORM 50
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Board of Trustees of the Leland Stanford Junior University, c/o Office of Research Administrator, Encina Hall, Stanford, Calif. 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, Human Resources Research Office, Arlington, Va. 22209 (Sub-Contract of Decisions & Designs, Inc., McLean, Virginia 22101)	12. REPORT DATE: OCTOBER 1977
	13. NUMBER OF PAGES: 126
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Engineering Psychology Programs, Code 455, Office of Naval Research, 800 No. Quincy Street, Arlington, Va. 22217	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release. Distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

A dissertation submitted to the Department of Engineering-Economic Systems and the Committee on Graduate Studies of Stanford University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, March 1977.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MULTI-ATTRIBUTE — PROXY — DECISION ANALYSIS

DECISION MAKING — FEASIBLE DIRECTIONS — DECISION THEORY

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In many decision problems, the possible outcomes have several important dimensions of value. To identify the optimal alternative, the decision analyst must assess the decision maker's preferences over these multi-attribute outcomes. Two rival procedures for solving multi-attribute preference problems currently exist. These two procedures, global → next page

(Continued)

Block 20 (Continued)

cont.

preference modeling and local preference modeling, each have advantages and disadvantages. This dissertation combines these two rival procedures in a new approach to multi-attribute decision making. This new combined method, called the proxy approach, uses the advantages of one technique to overcome the disadvantages of the other.

Global preference modeling procedures use normative assumptions together with a few assessments to construct a single function, in the large, ordering preferences over all outcomes. These global functions are mathematically simple and convenient, but they are very restrictive. The assumptions from which they are derived are reasonable locally, in the small, but when assumed globally, in the large, they often produce functions not truly representing the decision maker's preferences.

Local procedures provide an alternative approach that avoids restrictive assumptions. Instead of constructing a single preference function in the large, these procedures build a sequence of local preference models, in the small, each generating a trial solution. Each trial solution is better than its predecessor, so the trial sequence eventually reaches the optimum. Currently existing local procedures use successive linear approximations; these linear functions are poor preference models, so the iterative procedure is slow and inefficient. Since each iteration requires a time-consuming interaction with the decision maker, the slowly converging procedure is not practical.

This dissertation combines the desirable features of the global and local techniques in a new improved method. The normatively motivated preference models of the global procedure are incorporated as proxy functions in a local procedure. These proxies are better models of the true objective than are the linear approximations, so the resulting trial sequence reaches the optimum much faster. The new proxy approach yields rapid convergence without restrictive assumptions.

After the theoretical aspects of the proxy approach are developed, the new algorithm is applied to a curriculum planning problem. This practical application was successful; the decision maker, previously unfamiliar with decision analysis, was able to provide the assessments required at each iteration. With the help of various consistency tests, the tradeoff assessments generated trial solutions that converged rapidly to the optimal solution. Numerous insights into the interactive use of the algorithm were gained from this practical application.